

Nation Building in the Andes.
Evidence from the Chilean Invasion of Peru in the XIX Century

César A. Huaroto D.

January 2023

Abstract^{1*}

Can war increase nation-building and social capital? This paper studies the effect of fighting a foreign invasion via self-organized resistance on nation-building in a developing country during the nineteenth century. I use the case of the Sierra Campaign, the last stage of the Pacific War (1879-1884, between Peru and Chile), where local indigenous communities fought a guerrilla war against a professional invader army. Using a novel dataset on the events of the Campaign, I find that people living today in municipalities more exposed to this Campaign reduced indigenous ethnic self-identification but increased political participation, democratic values, and civic capital. To address endogeneity concerns, I controlled for geographical, climatic, and pre-conflict socioeconomic characteristics, tested robustness to matching algorithms, and estimated the Oster test, all indicating that the evidence is causal.

JEL Codes: D74, D91, H56, J15, O15, P16, Z21

^{1*} E-mail: cahuaroto@uc.cl. I want to thank my PhD advisor Francisco Gallego for his guidance and support. For their comments and suggestions, I want to thank Emilio Depetris-Chauvin, Felipe Valencia, Felipe Gonzalez, Miriam Artiles, James Feigenbaum, Laura Schechter, Jeanne Lafortune, Claudia Martinez, German Vega, as well as the audiences from BCDE2022; RIDGE EH Workshop; RIDGE Towards Sustainable Development; and internal seminars at the University of Piura (UDEP), Catholic University of Chile (PUC), and the University of the Pacific (UP). Fredy Gonzales, Alexander Pacheco, and David Zegarra provided superb research assistance. All remaining errors are mine.

1. Introduction

There is ample evidence of the role of war in fostering nation-building, which can be summarized in the well-known phrase of sociologist Charles Tilly that "war made the State, and the State made war" (Tilly, 1975). A large body of the previous literature has focused on the role of war in strengthening states by increasing their fiscal capacity, bureaucracy quality, and provision of public goods (Alesina et al., 2021; Becker et al., 2020; Dincecco & Onorato, 2018). This same literature argues that state-capacity-building efforts can later translate to increasing nation-building (measured as more significant identification with their nation-state) through different bundled mechanisms.

In this paper, I explore a novel mechanism: the role of the shared experience of facing a foreign military invasion using self-organized resistance in nation-building. In this scenario, the cohesive experience of facing a common enemy was the driver, not state policies. In the previous research, nation-building resulted from this shared experience and the need to build stronger national states that could afford the costs of war. States, then, needed to provide public goods to increase national identification from their population, so they could support the nation through the efforts that war demanded. In this paper, I aim to show that war can increase nation-building, even without state actions, an idea that the previous literature has not explored.

I analyze a guerrilla reaction during the Chilean invasion of the Peruvian Andes during the final stage of the War of the Pacific (1879-1884). This stage followed the occupation of Lima, the capital, and the Chilean army's defeat of its entire military. This locally organized guerilla response was known as the "Sierra Campaign" (SC) and was primarily driven by the local indigenous population. The resistance army could take advantage of the Andean geography to put up a tenacious defense against the occupying forces. After two years, the SC ended in 1884, and the Peruvian government signed the peace agreement soon after. Although the SC were defeated in the end, the local recollection of this event sees it as a victory and became an essential part of their culture (as documented in Sater, 2007; Manrique, 1981; Mallon, 1995, and Ministerio de Guerra, 1981a).

Like many developing countries, Peru has a significant prevalence of indigenous populations that were historically exposed to extractive institutions by colonialism. The combination of pre-colonial ethnic divisions and extractive colonial institutions has been

known as a cause of social disintegration (Nunn & Wantchekon, 2011), social unrest (Huaroto & Gallego, 2022; Guardado, 2018), low political inclusion (Lowes et al., 2017), and, more broadly, economic underdevelopment (Dell, 2010; Acemoglu et al. 2002). Therefore, the role of the democratic inclusion of the indigenous population, as well as the reduction of ethnic divisions, is of paramount importance for sustainable economic development, and it's scarce in the literature.

I reconstruct the list of events using the Peruvian army's official history books and then define treated units as municipalities with at least one relevant SC event within their area (battles, skirmishes, or the presence of an army headquarters, among others). I then estimate cross-section regressions, including pre-war socioeconomic (literacy, population density, ethnic composition, gender composition, among others), climatic (average precipitation and temperature), and geographic (altitude, slope, land suitability, and distance to rivers) conditions as covariates to address any potential selection bias.

The main findings of this paper are that people currently living in areas that were more heavily involved in the resistance have lower indigenous identification,² higher political participation,³ and higher civic capital population-wide.⁴ I then argue that this could explain why these areas currently have better economic conditions,⁵ giving additional evidence for the critical role of nation-building in long-term economic development in low-income countries, such as Peru.

To address endogeneity concerns, I performed different robustness checks, such as three different matching algorithms,⁶ finding that my main results hold. Considering the delta estimate of Oster (2019), I find that unobserved heterogeneity is unlikely to drive my main results. Also, I test that my main results hold even if controlling using "bad controls,"⁷ giving suggestive evidence that the change in identity beliefs is not a

² Specifically, I find that they report feeling less identified with an indigenous group (*Quechua, Aymara, or another native indigenous group*) while reporting the opposite towards a mixed-raced group (*mestizo, cholo, mulato, trigueño*). They are also less likely to speak a native language.

³ I find higher levels of political participation in four national presidential elections (1933, 1980, 2001, and 2021) in treated municipalities.

⁴ Children from treated municipalities show better academic performance in the subject "Civic Education" in a national standardized evaluation. People report more knowledge about democracy and its importance in opinion surveys.

⁵ I find that they had better quality housing in 1961 and 2017 in population censuses and higher income per capita in a representative household survey in 2018.

⁶ I estimate the same regressions using three different sub-samples, corresponding to three different matching algorithms: Mahalanobis Nearest Neighbor, Geographic Neighbor, and Entropy Balancing.

⁷ I use housing quality and income per capita in the household as bad control.

consequence of any other potential mechanisms. Finally, my results hold when using alternative, more conservative methods to measure standard errors, such as randomization inference and Conley's (1999) correction for spatial correlation, to account for the relatively small number of clusters and spatial correlation, respectively.

This paper contributes to three different strands of the literature. First, it builds on the extensive literature on the forces behind nation-building, which range from education (Alesina et al., 2021; Bandeira et al., 2018; Blanc & Kubo, 2021; Mitrinen, 2021), the influence of political leaders (Assouad, 2021), inter-ethnic contact (Bazzi et al., 2019), to entertainment media (Esposito et al., 2021). In particular, this paper is closely related to local ethnic identification as a barrier to nation-building (Depetris-Chauvin et al., 2020, Lupu & Peisakhin, 2017, Blouin & Mukand, 2019; Fouka, 2020; Marciante, 2022; Ronconi & Ramos-Toro, 2022).

Second, it contributes to the abundant literature on war's economic and cultural long-term effects (Feigenbaum et al., 2022; Grosjean, 2014; Riaño & Valencia, 2021; Bauer et al., 2016; Rohner et al., 2013; Blattman & Miguel, 2010). It speaks explicitly to work on war's effect on nation-building, which is heavily focused on its role in state capacity building (Alesina et al., 2021; Becker et al., 2020; Dincecco & Onorato, 2018; Gennaioli & Voth, 2015; Dincecco & Prado, 2012; Acemoglu et al., 2011; Mangini & Pefroff, 2022). A subset of this literature relates warfare events to nation-building, excluding the state's role (Dell & Querubin, 2018; Dehdary & Gehring, 2022; Alesina et al., 2020; Tur-Prats & Valencia, 2020; Heldring et al., 2022), although in most cases they focus on the potential adverse effects of forced assimilation policies.

The closest studies to this paper are Gehring (2021) and Steiner et al. (2022), which examine the short-run effect on self-identification from the Russian invasion of Ukraine in 2014 and 2021, respectively. Ochsner & Roesel's (2019) study is similar since it relates to the long-term impact of a past military invasion on the memory of the affected population. But the proposed mechanism is not associated with higher attachment to their nation but with grievances against the invaders. In contrast, I deal with the role of the shared experience of resisting a foreign invasion, a topic that has not been studied. Fontana et al. (2017)'s study is related but focuses on the political consequences of resisting an invasion, not on ethnic identification. Notably, these three papers deal with

the short-term effects of military invasions, while this paper presents evidence on the long-run effects.

Third, it contributes to the now vast literature on the causes of cultural persistence (Giuliano & Nunn, 2021; Lowes & Montero, 2021a, 2021b; Dell et al., 2018; Alsan & Wanamaker, 2018; Teso, 2018; Lowes et al., 2017; Jha, 2013; Voightlander & Voth, 2012; Nunn & Wantchekon, 2011). In particular, it relates to the subset related to identity formation (Bonomi et al., 2021; Kranton, 2016; Akerlof & Kranton, 2000; Benjamin et al., 2010; Shayo, 2009). This paper deals with ethnic identity, an understudied topic in Latin America but is highly relevant due to the significant indigenous populations in many countries that do not feel a strong attachment to their nation. This paper also relates to the subset of the literature that explores the cultural assimilation of migrants, which are similar to the indigenous population because they are ethnically different from the dominant group, and their social inclusion has significant economic consequences in the long run (Abramitzky et al. 2020; Advani & Reich 2015; Bleakley & Chin 2010).

2. Historical Background and Context

2.1 The War of the Pacific⁸

The War of the Pacific was an interstate conflict between Chile and a Peruvian-Bolivian alliance between 1879 and 1884. The war was also known as the "Saltpeter War" because control of this nitrate spurred the conflict. Saltpeter was used as fertilizer and to make explosives.⁹ The largest saltpeter deposits were in the Tarapaca and the Antofagasta regions in Peru and Bolivia, respectively. Due to the country's geographical proximity and better economic conditions, Chilean firms and workers were the first to mine these deposits in 1866. Then to protect the investors, the Chilean government made the Bolivian government agree to a stable 10-cent tax per 100 kg extracted in 1874. The war started because the Bolivian government doubled taxes in 1878. Chilean companies refused to pay based on the claim that these impositions were arbitrary and illegal considering the 1874 treaty, which resulted in the confiscation of their assets by the Bolivian government.

⁸ A comprehensive review of the history of this war can be found in Sater (2007).

⁹ By the end of the nineteenth century, its demand increased dramatically due to the expansion of the US into the West Coast, where it was relatively easy to transport due to access to the Pacific Ocean.

In response, the Chilean army occupied Antofagasta in February 1879, and the Bolivian government declared war the following month. Notably, this occupation faced little resistance since most of the desertic region of Antofagasta was virtually deserted outside of the Chilean miners. Peru then declared war on Chile in April due to a secret Bolivian-Peruvian treaty of mutual assistance in the case of war.¹⁰

At the end of the nineteenth century, most frontiers were not demarked, and Chile was increasing its boundaries to the south in the *Occupation of Araucania*.¹¹ Thus, it was reasonable for the Bolivian and Peruvian governments to be wary of Chilean ambitions and to expect that it would only be a matter of time before a large-scale conflict. This expectation can explain why there was a treaty of mutual assistance between them in the first place and why Peru involved itself so quickly in the Bolivian conflict. Finally, saltpeter access was critical since it was one of the few profitable natural resources available at the time (others, such as mining, required large amounts of difficult-to-find investment).

The events of the war can be seen in Figure 1. By 1880, all military power of Peru and Bolivia was defeated, and the latter surrendered. Peru did not resign, so the Chilean government decided to take Lima and force the Peruvian government to accept the peace treaty. Chilean troops landed close to Lima in January 1881 and overcame the city's defenses. The occupation of Lima ended three years later at the end of the war. The reasons for the length of the occupation were threefold. The first is that after the fall of Lima, the Chilean government assumed that, without a navy or army, Peru would have to accept their peace treaty, and because of this, it sent almost two-thirds of its troops back to Chile. The second reason was the chaotic internal Peruvian politics in Lima, which is discussed in the following subsection. Third, and arguably the most important reason, was the surprisingly tenacious local resistance in the central Andes under the leadership of General Andres Avelino Caceres in what is known as the Sierra Campaign (SC), which is detailed in Sub-Section 2.3 and is of primary interest to this paper.

¹⁰ However, the tax is considered by historians to be only the trigger of an inevitable conflict, just as the killing of Archduke Franz Ferdinand was the spark that lit the tinderbox of World War I. Only by the end of the nineteenth century did South American countries have enough economic and logistical resources to engage in warfare against each other. Also, the chaotic independence processes in the first half of the century left many boundary controversies between the new nations.

¹¹ The Occupation of Araucania was a series of military campaigns between 1861 and 1880 by the Chilean army and settlers in *Mapuche's* indigenous territory, which led to the incorporation of the Araucania region into Chile's national territory.

HERE FIGURE 1

2.2 *The Political Chaos in Peru*

Like many Latin American countries at the time, Peru had a weak democracy and around half-century since it was emancipated from colonial rule. At the beginning of the war, the president was Mariano Prado. He fled to Panama in December 1879 after the loss of the *Land Campaign*, fearing for his life due to profound disapproval among the population of how he managed the conflict.

Soon after, Nicolas de Pierola, a typical *caudillo*,¹² took advantage of the situation, initiated a coup, and then refused to sign a peace agreement claiming it was still possible to recover the lost territories. He fled the city in January 1881 when the Chilean army defeated Lima's defenses.

The Liman elite appointed a new emergency president in March 1881, Francisco García Calderón. Calderon's regime was rapidly accepted by Chileans, who even provided him with a small army and funds, hoping he would acquiesce to their demands. Surprisingly, he refused to do so, and the Chileans exiled him to Santiago, the capital of Chile, in November 1881. Before his exile, he appointed Admiral Lizardo Montero, a respected military leader, as the country's president. This thwarted Chilean peace treaty attempts since Montero was in Arequipa, far from the occupation.

The situation became more complex for Chileans since, as I discussed in the following subsection, a resistance army emerged in the Central Andes. This also added complications for any political leader in Peru to accept a peace agreement while there was still active warfare that they couldn't control. Thus it was only after the defeat of this resistance that a peace agreement was made. The war ended with the peace treaty of Ancon in October 1883, only one month after Caceres's defeat in the Huamachuco battle.

2.3 *The Sierra Campaign*

After the occupation of Lima, Chileans did not expect additional resistance. However, General Andres Avelino Caceres built a new army in the central Andes, claiming that Peru could recover the territories lost during the war. While they were somewhat

¹² *Caudillos* is the name dictators have in Latin American countries during the first 50 years after independence. They also typically promise populist policies.

unrealistic, the Chileans could not dismiss these aspirations since they threatened to extend the occupation, which was expensive in terms of monetary and personnel resources. The events of the SC can be seen in Figure 2.

Caceres was a native of the central highlands, a descendant of landlords who had learned to speak Quechua, the local indigenous language, to communicate with the estate's workers. Although he was not ethnically indigenous, he claimed he was a descendant of Catalina Huanca, a princess of an ethnic group in the Central Andes, the Wankas. Under his leadership, the resistance gathered support from the local indigenous population. It took advantage of the geographic conditions in the Andes to engage in guerrilla warfare, to which the Chilean military was vulnerable. This last stage of the war is regularly considered to be the most violent and exhausting. According to Sater (2007), this was a "widespread, prolonged, brutal, and eventually futile guerrilla campaign," lasting nearly three years between February 1881 and October 1883.

A notable feature of the Sierra Campaign (SC) is that it only occurred in the central highlands. The mountainous conditions benefited guerrilla warfare over the more traditional fighting where Chile dominated. The central Andes were also reasonably close to occupied Lima; thus, the resistance had an opportunity to engage.

Naturally, after the Campaign's start, Chile's main objective was to defeat Caceres and extinguish the remaining resistance so that they could pressure the political and military leaders, as well as public opinion, to accept the peace agreement. But Caceres proved a formidable rival and defeated the Chileans on several occasions. Although Caceres was finally defeated in July 1883 in the battle of Huamachuco, he managed to avoid being captured and even attempted to continue the resistance.

Due to the enormous losses of his military forces, the exhaustion of the economic resources to sustain the resistance, and the support of the national elites on the peace agreement, continuing with the resistance proved impossible. Peru signed the peace treaty a few months later, in October 1883.

HERE FIGURE 2

3. Data Sources and Description

This study uses a variety of data sources. In the following sub-sections, I will briefly discuss the origin of each data source and the type of information used. I will also present details on the variables' construction and disaggregation levels. Descriptive statistics and the source of all variables are given in Table 1 in order of appearance in the paper.

HERE TABLE 1

3.1 Reconstruction of District Boundaries of 1879

There were 1874 districts in the last Peruvian census in 2017, while in the 1876 census, there were only 757 districts (from now on, these will be called historic districts). I thus adapted the current data from modern district boundaries to historical ones. Specifically, I identify to which historical district every modern district belongs using the creation laws and the list of localities (towns, villages, etc.) for each 1876 district to map them onto the current landscape.

The analysis excludes districts in Lima (the capital of the country) and the region of Tacna (occupied by Chileans between 1881 and 1929). Naturally, I excluded the provinces of Arica and Tarapaca, which became Chilean after the war, and all the Amazonian regions because they were nearly or entirely unpopulated.¹³ The final sample includes 1496 present-day districts that belonged to 638 historic districts.

3.2 Exposure to the Sierra Campaign (1881-1883)

To construct a measure of exposure to the SC, I use the documentation collected by Ministerio de Guerra (1981a, 1981b, 1981c), which included all events of the SC, giving information on its timing, location, and a brief description of each event (number of casualties, actors involved, etc.)

According to this documentation, between January 1881 and July 1883, there were 199 resistance events against Chilean occupation. I excluded 104 of them for four reasons. First are those for which it was impossible to plot on a map or were in areas not included in the analysis. Second, I exclude events if they were minor or nonviolent (e.g., a political

¹³ Also, historical data from the 1876 census was missing for the provinces of Huanta in Ayacucho and Anta in Cuzco. Thus, those districts are also omitted.

discourse, march, or peaceful protests). Third, I excluded events after the decisive Huamachuco battle, those outside the central Andes, if they were irrelevant in size or spontaneous riots or protests unrelated to the resistance led by Caceres.

This leaves 95 events. All these events were either violent encounters between the guerillas and Chileans or were the location of any army's headquarters.¹⁴ I then mapped this event-level data to the municipal level and created a dummy variable since most municipalities had only one event. After this process, I ended up with a treated group of 37 historical districts exposed to at least one SC event. Figure 3 presents the borders of treated districts and the locations of excluded events.

HERE FIGURE 3

In Figure 4, I present three maps to give more information on the treatment districts' locations. In the left panel, the country's topography is given, showing the notable differences in altitude caused by the Andes. The central panel shows three groups: in orange, the 37 treated districts; in light blue, the 601 non-treated districts included in the study; in grey, the areas of the country that were not included at all. Finally, in the right panel is a zoom into the country's central region, where most SC events happened.

HERE FIGURE 4

3.3 Control Variables

An extensive set of control variables are used to control for time-invariant characteristics that may affect the treatment and the outcomes. They come from geographic, satellite, historical, and present-day demographic characteristics, all of which are presented below in detail. Descriptive statistics of these variables are shown in Table 1 at the end of this section. All regressions include them as controls; they are also used for the matching robustness checks.

¹⁴ Even without a violent encounter, the location of a Chilean army headquarters can be considered a violent experience for the local population due to related abuses or repression. Meanwhile, a resistance's headquarters can be interpreted as a sign of the strong commitment of the population to the resistance.

A. Geographic and Satellite Data

I use different types of geographic satellite data, but I estimate the average elevation, slope, and historical rainfall data with ArcGIS for all districts¹⁵. The former two types of data come from NASA's STRM, the latter from the Climatic Research Unit of East Anglia University. Finally, I also include data from the Land Suitability Index from Ramankutty et al. (2001); this index takes a value within the range of 1 and 0, being 1 for land that is very good for agriculture and 0 if it's infertile.

B. Historical Data

I obtained data from the 1876 national population census from official publications (MGPO, 1878) that summarized information at the district level, which I digitized. The dataset I constructed is the only pre-conflict one available and is primarily used for testing for any previous differences between treatment and control groups. It includes population numbers, gender, ethnic composition, religion, literacy rates, nationality, rurality, and the percentage of people living in rural estates, called *haciendas*.¹⁶

In addition, I collected colonial and pre-colonial data to control for any historical confounding factors. Specifically, I include a dummy variable for if the district was subjected to the Mining Mita during colonial times (Huaroto & Gallego, 2022).¹⁷ Also included as controls are the population density and the percentage of cropland in 1500, estimated using data from Golderwijk et al. (2010).

C. Present-day Control Variables

For the individual-level data, when available, I include as control variables age, gender, education level, and if they live in a rural area.

¹⁵ In specific, I use population weighted averages. Since I have geographic data of these variables at the grid level, I use the population living in each grid to weight it. This allows to avoid using geographic characteristics of unpopulated areas into account and to give more importance to in the average to the largest cities.

¹⁶ The share of the population living in haciendas is a proxy of the importance of haciendas in the districts, which was an important feature of rural Peru at the time. The larger the importance, the more inequal the relation between landlords and the indigenous population.

¹⁷ The Mining Mita was a forced labor and migration institution that was implemented during the Spanish colonization of Peru and Bolivia which created a forced labor for the Potosi and Huacavelica mines. It's important in the context of this since this colonial institution was focused on the country's highlands. Although it was eliminated in 1811, it has been proven to have lasting effects on economic development and social unrest (Dell, 2010; Huaroto & Gallego, 2022).

3.4 Outcome Variables:

A. Identity and Ethnicity

My primary outcome of interest is a measure of ethnic identification at the individual level gathered by the 2017 National Population Census and the 2017 National Household Survey (ENAH), a nationally representative survey similar to many others in developing countries. The first gave 12,175,049 observations, the total population above five years old in the 638 historical districts mentioned in Section 3.2. The second provided 53,827 observations, who are above 14 years old.

Both sources ask respondents to select their primary (if any) ethnic identification. Specifically, the question is: "Based on your customs and ancestors: To which group of the following do you feel or consider yourself most related?". The alternatives were: i) Quechua, ii) Aymara, iii) Amazonian native/indigenous, iv) Another native or indigenous group, v) Black, brown, mulatto, or any other type of afro Peruvian group, vi) White, vii) Mestizo (white and indigenous mixed blood), and viii) Other.

I then construct four binary measures. The first would take the value of one if the respondent selected any of the first four options, which I interpret as having an indigenous ethnicity. The other variables take the value of one depending on the option chosen, which I then assign as "white," "black," or "mestizo," respectively.

The second set of identity and ethnicity variables is the population percentage by the district in terms of the first language in 1961, 1981, and 2017. Although this indicator is not a precise measure of ethnic self-identity, it is often correlated with it and has the advantage of being historically available. In this case, the question was: "Which language did you learn during childhood?" Similarly to the self-identity question, the alternatives were Quechua, Aymara, other native languages (Ashaninka, Awajun, etc.), Spanish, and other foreign languages (Portuguese, English, etc.). In this case, I estimate the percentage of the population in the district that learned Spanish, a Native language (which includes all native languages in Peru), and other languages.

The last set of ethnicity variables comes from the historical censuses of 1940 and 1876. Respondents were not asked about their ethnicity or first language in those censuses. Instead, the surveyor assigned an ethnic group to the respondents by appearance. This

difference allows me to rule out a potential confounder: race. Since it's an external person who is in charge of determining ethnicity, it's possible to evaluate if there are visible differences in the tone of the skin, which might be related to a differential racial composition. In this case, the variable is the percentage of the population by the district classified as indigenous.

Data for 1981 and 2017 was estimated using individual-level data, while for 1961 was collected and digitalized district-level data from DNEC (1965). Similarly, data for the 1940 and 1876 censuses were collected and digitalized from MHC (1945) and MGPO (1878). To make results comparable with individual-level data used for ethnic self-identification, I use the district population for that year as a sample weight.

B. Electoral Outcomes and measures of Democratic Capital

I use the number of emitted votes in the district as a proxy for political participation. I collected data on this indicator for four presidential elections: 1933, 1980, 2001, and 2021. I choose to use the number of casted votes in the district since there are no reliable measures of eligible voters in the two first waves of data. Also, in the first, the election was restricted to literate men, which made comparisons between elections complex. To avoid the potential problem of the district population, I use the district's people in the closest census as a control variable (1940, 1981, 2007, and 2017, respectively).

Additionally, to better control for geographic and economic differences, I include district literacy rates as a control in addition to all the control variables mentioned above. Data for the elections of 1933 and 1980 comes from DNE (1933) and JNE (1982), respectively. Data for the elections of 2001 and 2021 come from the official webpage of the National Electoral Office (ONPE).

The second measure of this section comes from the 2018 Student National Assessment Test (ECE) math, reading, and civic education scores collected by the Education Ministry. The math and reading tests cover a representative sample of children in second grade (5,433 seven-year-old children), while the civic education test is a sample of sixth-grade students (5,200 eleven-year-old children). Data is at the individual level, so I include student gender and school-level controls (private/public, urban/rural).

The third measure comes from the special opinion module on "Governability, Democracy, and Transparency" from the 2017 National Household Survey (ENAH). This module has a sample of 12,844 individuals, randomly selected adult members of each household surveyed. I take advantage of this unusual inclusion of the political opinions module since it has a large sample size.¹⁸

From this module, I use three questions. The first is: "Do you know what democracy is?" with a "Yes/No" option. I then constructed a binary variable, which takes the value of one if the respondent answered positively.

Those who say no to the first question do not get asked the following questions (around two-thirds). The second is a multiple-answer question: "Select all the alternatives that indicate what democracy is important for?" the alternatives are five: i) Elect authorities, ii) Be represented, iii) Achieve family wellbeing, iv) Make people's rights respected, and v) For nothing. I then construct five binary measures, each taking the value of 1 if it selects each alternative.

The third relates to the primary outcome of the paper: group identity. Specifically, and the single-option question is: "To which group do you feel more identified?" to which the alternatives are: i) Region, province, district or city, ii) Ethnic group or race, iii) Indigenous peasant community, iv) Religious group or v) Others. Similarly to previous questions, I then constructed five binary measures, each taking the value of one if the respondent selected each alternative.

C. Economic Development

The final set of outcome variables is related to long-term economics. The first is literacy rate, for which I used the historical population censuses of 1940 and 1961 from MHC (1945) and DNEC (1965), respectively. I then complemented this data with literacy estimations at the district level using individual-level data from the 1981 and 1993 censuses. I also take advantage of the fact that literacy rate numbers were available in 1876, three years before the war, so it's possible to control for pre-treatment differences.

¹⁸ In contrast, specialized opinion surveys, such as LAPOP, World Values Survey, and Latinobarometer, which include questions related to democratic values, lack the sample size needed to study the effect of a localized event, such as the SC. For instance, LAPOP 2017 survey, the more prominent in sample size among them, had only 2,647 observations for the complete country.

As previously, I make results comparable with individual-level using the population in the district in that year as a sample weight.

Finally, I estimate different long-term economic development proxies using census data. The first four are related to housing quality in 1961 (percentage of households in the district with good-quality walls, roofs, and floors) and 2017 (a dummy indicating that the individual lives in a dirt-floor house). The final proxy is the per capita income obtained from the 2017 ENAHO survey.

4. Identification Strategy

The main estimating equation using individual-level data is:

Long-run effect estimator for individual data: $y_{ij} = X_{ij}'\varphi + \gamma T_i + \omega_{ij} \quad (1)$

where the outcome variable y_{ij} , X_{ij} and the error term ω_{ij} are at the level of individual i living in district j . In this estimation, I report standard errors clustered at the district level. Where T_i represents a dummy variable that takes the value of 1 if district municipality i was exposed to the SC. The X_{ij} term represents a set of geographic controls, pre-war socioeconomic conditions at the district level, and demographic controls at the individual level (gender, age, etc.). ω_{ij} represents an i.i.d. error term. Here the coefficient γ is our estimate of the SC's long-run effect. This specification will be used with microdata from surveys or censuses.

The main estimating equation using district-level data is:

Long-run effect estimator for district data: $y_i = X_i'\varphi + \gamma T_i + \varepsilon_i \quad (2)$

where the specification is identical to the previous, but without individual-level controls, and uses robust standard errors and population as an expansion factor to recover population-level estimates to make the results comparable with those from equation 1. Where T_i represents a dummy variable that takes the value of 1 if the district municipality i was exposed to the SC fighting. X_i term represents a set of geographic controls and pre-war socioeconomic conditions at the district level. ε_i represents an i.i.d. error term.

My main estimation method for equations (1) and (2) is Ordinary Least Squares (OLS), where I assume that conditional on the control variables, γ identifies the causal effect of

being exposed to the SC on the outcome variables. As discussed below, I also present a set of robustness exercises to study how potential confounding effects related to unobservable variables can explain my OLS results. This will be discussed in more detail in section 8.

Finally, it was only possible for an outcome variable (literacy rate) to use a Difference in Difference (DID) specification. In this case, I estimate equation 3:

$$\text{DID estimator: } y_{it} = \alpha_i + Post_t + Post_t X_i' \varphi + \lambda Post_t T_i + v_{it} \quad (3)$$

Where $Post_t$ is a dummy that takes the value of one if the observation is in period t observed after 1883, the end of the SC, and is 0 otherwise. The variable y_{it} represents the outcome for district i in year t . X_i represents a set of geographic controls and pre-war socioeconomic conditions that are interacted with year fixed effects to control for any differential trend caused by these variables. The list of control variables is presented in Table 1. α_i is a vector of the district-level fixed effects. Here T_i represents a dummy variable for if district i was exposed to the SC. Here the coefficient λ is our estimate of the DID effect of the SC. This specification will be used for all the district-year data. Finally, v_{it} represents a two-way clustered standard error.

The main advantage of DID estimates over OLS is that it's possible to control for pre-treatment differences in the outcome variable. For identification, I rely on the assumption of parallel trends to identify causal effects.

5. Main Results

A. Effect on Ethnic Identity

Table 2 reports the OLS estimates of equation 1 for the SC's impact on ethnic identification. All regressions include district-level controls, such as pre-war socioeconomic, historical, geographic, and climatic conditions, as discussed in Section 3.3, and individual-level controls, such as age, gender, education level, and if the respondent lives in a rural area.

The first four columns present results using census data, and columns 5 to 8 with survey data. Columns 1 and 5 report the effect on the probability that the individual self-identifies

as indigenous, 2 and 6 do the same with white, 3 and 7 with mestizo¹⁹, and, finally, columns 4 and 8 for all other ethnic identities.

As seen in Table 2, for both census and survey, there is a positive effect of the SC on the probability of self-identifying as "mestizo" and an opposite effect on the likelihood of indigenous identities, such as *Quechua* or *Aymara*. No effects can be seen on the probability of choosing white or another ethnicity.

Peru, like many Latin American countries, had a large indigenous population before colonization and had not received large European or African migration waves. Most of its people are descendants of the original indigenous populations and Spanish colonizers. This can be seen in the control group mean, where 55% of the population self-identifies as mestizo and 31% as indigenous in the population census.

HERE TABLE 2

This effect on identity could be attributed to different causes related to the impact of the SC. In this paper, I argue that the effect is cultural, not associated with other economic or social changes. As evidence of this, in Table 3, I present the exact estimation of the previous Table but include proxies of economic wellbeing as "bad controls." For the census, I include a dummy indicating if the individual's house has a dirt floor, a commonly used proxy of poverty. For the ENAHO survey, I use the per capita income of the individual's household as a control. As shown in Table 3, results are unaltered by the inclusion of these controls, supporting the idea that economics is not affecting the relationship between SC exposition and identity beliefs.

HERE TABLE 3

Moreover, in Figure 5, I present the exact estimates for different sample subgroups. The first is by age (12 to 17, 18 to 39, 40 to 64, and 65 and above), the second is by the location type (urban or rural), the third is by education level (completed elementary or less, completed or some secondary school, and any advanced education), and fourth is by gender. As seen, for all sub-groups, there is a negative effect on the probability of self-identifying as indigenous. Its size is qualitatively similar, indicating that the point

¹⁹ Mestizo is a common term used in many LAC countries to define people with white and indigenous mixed blood.

estimate is spread along the population, reinforcing the idea of a cultural mechanism driving it.

HERE FIGURE 5

B. Testing the Cultural Mechanism: Evidence from Migrants

I propose two additional tests for this hypothesis. The first is to focus on immigrants in SC districts. If the cultural effect exists, migrants from untreated districts would not have the same beliefs as natives. In Table 4, I present evidence supporting that claim. I replicate the same analysis in Table 2 but divide people in treated districts into two groups: natives and migrants. The effect is only present for natives, and I find no statistical differences between migrants and people living in untreated districts.

HERE TABLE 4

The second test is the opposite. If culture drives identity, people born and raised in treated districts should carry these same beliefs when they migrate to untreated districts. In this case, I focus only on migrants living in the city of Lima, the national capital, as it is the leading destination for migrants. As seen in Table 5, when I replicate the analysis using only migrants (both from treated and control districts) living in Lima city, I find similar results but smaller magnitude.

HERE TABLE 5

C. Using alternative measures of ethnicity

So far, I have presented evidence that people in SC-affected areas are more likely to identify as mestizo than as indigenous. In this subsection, I provide two additional measures of ethnicity to study the hypothesis and the previous results in more detail.

The first is to explore not a belief but rather a cultural tradition, specifically learning an indigenous language. Although it is not a precise measure of identity, it is undoubtedly correlated. It can be interpreted as a more extreme measure of identity since preserving an indigenous language has social and economic costs for parents and communities but is also seen as critical in preserving indigenous cultures. Of course, some people can still identify as indigenous but not speak the given indigenous language. However, it's harder

to find individuals that speak a native language who do not identify themselves as indigenous. An additional advantage of this measure is that I have longitudinal data from 1961, 1981, and 2017. Thus, it is unlikely it is due to any random outside factors.

In Table 6, I report the difference in the probability of knowing an indigenous language as its first language. As seen, the impact of the SC is statistically significant for all three years: people living in treated districts have a lower probability of an indigenous language being their first language and a higher probability of their first language being Spanish. Something worth noting is that the effect became significantly more substantial in 1981 and 2017, concerning 1961. At the same time, the percentage gets lower in the control group, indicating that in relative terms, the difference has become even more prominent over time.

HERE TABLE 6

The second is to rule out an alternative, although credible, explanation. These cultural differences could be related to a demographic change caused by the SC. First, the war could have disproportionately killed indigenous people, thus having a long-term change in the population. Second, it could be that the SC created higher white immigration from white people or reduced out-migration, which could also impact ethnic identification. Third, it might also be that Chilean soldiers, primarily white, might, forcibly or voluntarily, impregnate local women and, in that manner, have modified ethnic composition, too.

The problem with these alternative explanations is that they would demand evidence of current racial demography. As far as I am aware, there are no reliable measurements of racial composition in the country, even by ancestry (e.g., skin color). Moreover, in Peru, where the racial mixture has occurred since early colonization, it's virtually impossible to distinguish mestizos from indigenous people phenotypically. Interestingly, in the first two censuses of the country (in 1876 and 1940), there was an attempt to measure race by appearance, which the enumerator assigned. In other words, there is a rough, although very subjective, a measure of race that is potentially interesting for those years. Results are presented in Table 7, and there are no statistical differences elsewhere, suggesting that potential racial differences aren't a factor.

HERE TABLE 7

6. Effect on Electoral outcomes and Civic Capital

In this section, I will discuss another legacy of the Sierra Campaign: It increased civic capital. For this, I rely on data on participation in the presidential elections of 1933, 1981, 2001, and 2021.

An important caveat about the 1933 election is that rules for participation were markedly different from the other three. Only literate men older than 18 were allowed to vote, and voting was voluntary among them. In contrast, voting was mandatory in the other three elections. Punishments for not voting included a fine and some administrative penalties until the fine it's paid (not being able to marry, sign a contract with a notary, or take a government job, among others). Another issue with comparing over time is there isn't any good information on the total number of voters for the 1933 and 1981 elections. These two reasons made using the turnout rate complicated, and thus I decided to use the number of votes in each election as the outcome variable. To better control for potential confounders related to the size of the district and the accessibility of voting places (located mainly in cities), I included as additional controls the population (obtained from the most recent census), the rural population, and the total literate population.

In Table 8, I present my results, and as can be seen, being in a treated district is related to increased voting rates. For the first two elections (1933 and 1981), the SC increased the number of votes by 19%; for the 2001 and 2021 elections, the effect is between 6 and 10% more. To put this estimation in contrast, on average, the number of casted votes increased by 30% between 2001 and 2021; therefore, the SC increased this number of votes between 10 and 6% in 2001 and 2021, respectively, indicating that the size of the SC is relevant.²⁰

HERE TABLE 8

Now, I turn to a different measure of political engagement: civic education. I use the results of a National Standardized Evaluation (ECE) that elementary students took in

²⁰ In Appendix Table A.1, I present the effect on turnout rate and the rate of valid votes among casted votes for the 2021 presidential election and the 2018 municipal election. The main results hold that it's possible to see that in the presidential election of 2021, there was a 2.2 pp. higher turnout rate and an increase of 3 pp. in the valid vote rate (share of valid votes among casted votes) in the presidential election, and even larger for the municipal election (around 9 pp. and 8 pp. for each, respectively). As mentioned before, it's only possible to estimate the effects on the turnout and valid vote rates for 2001 and 2021.

2018. This evaluation included three subjects: reading and math for second-grade and sixth-grade civic education. This measure complements the previous election data well since it provides evidence that the results are not driven by political factors in the district (higher budgets or higher political competition, among others) but are related to higher local interest in politics or civic values.

In Table 9, I use the standardized scores in all three subjects. Columns 1, 2, and 3 show the differences between treated and control districts in civic education, math, and reading, respectively. They are only statistically significantly different for civic education; students in treated areas are notably better in this subject. The result is aligned with all the previous evidence shown in the paper, suggesting that the primary impact mechanism of the SC is related to cultural values. These results indicate that differences in the supply of education could not explain previous effects; if that were the case, we would be able to find the impact on the other two subjects.

HERE TABLE 9

Finally, I turn to my last piece of evidence, which links to ideas of identity and politics, as it surveys democratic values and identification with the country. Data comes from the "Governability, Democracy, and Transparency" module from the 2017 National Household Survey (ENAH). As mentioned previously, the larger sample size is the main advantage of using this data over other surveys such as LAPOP or Latinobarometer.

In Table 10, I present evidence of differences in beliefs about democracy. In this case, I also estimate the SC's differential effect on the more rural district in 1876 (using the percentage of the rural population among the total population). As seen in column 1 of Table 10, I show that individuals in treated districts better understand democracy, but only in rural areas. Similarly, in columns 2 and 3, I find that in treated rural areas, people are more likely to select two options when asked about the functions of democracy. Those two options are "to choose authorities" and "to be represented," indicating that they value democracy more (or at least consider it more important for their wellbeing).

Finally, in columns 4 and 5, I also show that they are more likely to report feeling increased identification towards the country (or another administrative unit within) than

with the local peasant community.²¹ Since these peasant communities are related to managing agricultural lands, it makes sense that the effect is concentrated in rural areas.

HERE TABLE 10

7. Effect on Economic Development

In this section, I present evidence that the SC increased economic development. The first proxy I use is the literacy rate for which I have historical data. Then I use other present proxies, such as housing quality and income.

In Table 11, I report a DID estimate on the effect of SC on literacy. I find statistically significant positive results for 1940, 1961, 1981, and 1993. Note I stopped in the 90s as literacy is only a good proxy when there is limited access to education. Notably, I had the opportunity to control for pre-existing differences in the outcome variable. Interestingly, the effect is even more significant in more rural districts, continuing with the previous result that the changes are more prevalent in these areas.

HERE TABLE 11

In Table 12, I report the OLS estimates of the SC's effect on different proxies of economic wellbeing. The first three columns present housing-quality outcomes from the 1961 census. I find statistically significant positive impacts for two: having a good-quality roof and floor. Column 4 uses microdata from the 2017 census and shows a negative effect on the probability of having dirt (low-quality) floors, indicating that the economic conditions continue to improve in treated districts. Finally, column 5 presents the impact on the household income per capita logarithm, finding that households in treated districts have higher incomes.

HERE TABLE 12

²¹ Peasant communities are prevalent in the rural areas of Peru and are associated with the colonial self-government institution called *ayllu*. To some extent, this has persisted to the present, much like similar institutions like the *ejidos* in Mexico.

8. Robustness Checks

8.1 Matching in observables

A potential concern is that there might be confounders. As shown in Figure A.1. in the Appendix, districts exposed to the SC were at higher altitudes, were more populated and less rural, and were more likely to have been exposed to the Mining Mita. Even after including these variables as controls, it is plausible that these confounders (or others) could be causing part of the estimated effect.

To address this, I use three different specifications for the control group. The first two are based on two commonly used matching algorithms: nearest neighbor matching and entropy balancing (Hainmueller, 2012). The third is restricting the control group to the exposed districts' geographic neighbors. Each specification has advantages and disadvantages, so using all three gives a broader examination of the result's robustness. In Figure A.2 in the Appendix, I show the distribution of the control group used in each case. Each potential control unit is assigned a weight for the entropy balancing matching algorithm, where observations with larger weights are more important for the comparison.

In Table A.2, in the Appendix, I show that the paper's main results, as presented in Table 2, are robust to selecting the control group using any of these matching algorithms. Although the estimated size varies, the coefficient signs are the same and statistically significant for all three algorithms.

8.2 Estimating the size of the selection in unobservables with Oster's (2019) methodology

Oster (2019) developed a methodology to measure the stability of regression coefficients, given the inclusion of control variables, to approximate any potential bias that unobservable factors could have. The main idea is that if the coefficient is stable and the model's predictive power increases with the inclusion of control variables, then there is a low risk that the inclusion of additional covariates (e.g., unobservables) would alter the results significantly.

With this method, it is possible to obtain a delta coefficient that measures how significant this threat is for a particular estimation. Notably, the minimum value recommended for

the delta coefficient when considering the estimate's robustness is below -1 when negative and above +1 when positive.

In Table A.3, in the Appendix, I test the robustness of the Table 2 results using the Oster (2019) methodology. In columns 1, 3, 5, and 7, the effect of the single treatment dummy is included, while columns 2, 4, 6, and 8 have all control variables. As can be seen, the estimated coefficients vary slightly while the R-squared increases notably. Each pair of columns presents the Oster (2019) delta coefficient estimation, showing the main results presented in Table 2 are robust. Thus these results are also unlikely to be driven by the selection of unobservables.

8.3 Robustness in the estimation of standard errors

Finally, due to the small size of the treatment group (only 37 districts) and the fact that they are geographically close, there is a concern that my estimations might be overestimated. Therefore, I estimate two alternative measures of standard errors: Randomization Inference (RI) and Conley's (1999) correction for spatially correlated errors. In Figure A.3, I show the coefficient distribution in a Monte Carlo scenario, assuming that the actual effect would be zero. As can be seen, the estimate is very far from the hypothetical distribution; thus, it is implausible that the results are driven by outliers or chance. Table A.4 presents the Table 2 results using Conley's standard errors for 50, 100, and 150 km spatial correlation; the main results hold for all of them.

9. Concluding Remarks

In this paper, I study the medium- and long-run consequences of exposure to the Sierra Campaign (SC) on ethnic self-identification in the Andes of Peru. The SC was the last stage of the War of the Pacific (1879-1884), where Chile fought and won against a Peru-Bolivian alliance. The Peruvian resistance ended in 1884, and the government signed the peace agreement soon after. Although they were, in the end, defeated, the areas involved in the SC see their particular part of the war as a victory.

My main result is that the SC positively affected nation-building by reducing ethnic self-identification and increasing political interest. Today, people in treated districts are more likely to identify with the larger and more inclusive mestizo ethnic group and less likely to identify with indigenous groups, reflecting social cohesion on a cultural level. They

are also more likely to vote, have a more substantial valuation of democracy, and hold higher civic capital. SC-affected districts also have better economic conditions both in the past and present, indicating that increased nation-building might have created better economic conditions in the long run.

Peru has a history of exploiting and excluding indigenous populations, like many developing countries. The combination of pre-colonial ethnic divisions and extractive colonial institutions has been known to cause social disintegration, social unrest, low political inclusion, and, more broadly, less economic development. In this context, a violent, yet cohesive shared experience, such as war, can help fragmented societies overcome these cultural issues and increase their cooperation and identification with the state, which might foster economic development in the long run.

This paper contributes to the longstanding literature that has explored the role of war on nation-building but in a context where the main channel is the shared experience of facing a foreign military invasion using self-organized resistance. In other words, while previous literature focuses on the role that the state had in both increasing national self-identification (replacing previous ethnic or regional ones), while also expanding its size (fiscal capacity and provision of public goods), in this paper, I present suggestive evidence that this national identification can emerge even without the role of the state.

Being forced to resist a powerful enemy can eliminate previous grievances. In the future, the remembrance of this event can be impactful in increasing national identification, and this could have, in the long run, a positive event that can overcome the economic and social short run costs of war. This can have important implications for ongoing conflicts, such as the ones occurring in sub-Saharan Africa, where ethnic fragmentation is also prevalent. But also in the case of Ukraine, where we see that even against a more powerful enemy, the local population can put a tenacious resistance that, even when costly today, can have significant positive consequences in the future.

10. References

- Abramitzky, R., Boustan, L. & Eriksson, K. (2020). Do Immigrants Assimilate More Slowly Today Than in the Past? *American Economic Review: Insights*, 2, 125-141.
- Acemoglu, D., Cantoni, D., Johnson, S. & Robinson, J. (2011). The Consequences of Radical Reform: The French Revolution. *American Economic Review*, 101, 3286-3307.
- Acemoglu, D., Johnson, S. & Robinson, J. (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *Quarterly Journal of Economics*, 118, 1231-1294.
- Advani, A. & Reich, B. (2015). *Melting Pot or Salad Bowl: The Formation of Heterogeneous Communities*. Institute for Fiscal Studies Working Paper W15/30.
- Akerlof, G. & Kranton, R. (2000). Economics and Identity. *Quarterly Journal of Economics*, 115, 715-753.
- Alesina, A., Giuliano, P. & Reich, B., (2021). Nation-building and Education. *Economic Journal*, 131, 2273-303.
- Alesina, A., Reich, B., & Riboni, A. (2020). Nation-building, nationalism, and wars. *Journal of Economic Growth*, 25, 381-430.
- Alsan, M. & Wanamaker, M. (2018). Tuskegee and the Health of Black Men. *Quarterly Journal of Economics*, 133, 407–455.
- Assouad, L. (2021). *Charismatic Leaders and Nation-Building*. Unpublished manuscript.
- Bandeira, O., Mohnen, M., Rasul, I., & Viarengo, M. (2018). Nation-Building Through Compulsory Schooling during the Age of Mass Migration. *Economic Journal*, 129, 62-109.
- Bauer, M., Blattman, C., Chytilová, J., Henrich, J., Miguel, E. & Mitts, T. (2016). Can War foster Cooperation?. *Journal of Economic Perspectives*, 30, 249:74.
- Bazzi, S., Gaduh, A., Rothenberg, A. & Wong, M. (2019). Unity in Diversity? How Intergroup Contact Can Foster Nation Building? *American Economic Review*, 101, 3978-4025.

Becker, S., Ferrara, A., Melander, E. & Pascali, L. (2020). *Wars, Taxation and Representation: Evidence from Five centuries of German History*. Unpublished manuscript.

Benjamin, D., Choi, J. & Strickland, A. (2010). Social Identity and Preferences. *American Economic Review* 100, 1913–1928.

Blanc, G. & Kubo, M. (2021). *Schools, Language, and Nations: Evidence from a Natural Experiment in France*. Unpublished Manuscript.

Blattman, C. & Miguel, E. (2010). Civil War. *Journal of Economic Literature*, 48, 3-57.

Bleakley, H., & Chin, A. (2010). Age at Arrival, English Proficiency, and Social Assimilation among US Immigrants. *American Economic Journal: Applied Economics*, 2, 165–92.

Blouin, A. & Mukand, S. (2019). Erasing Ethnicity? Propaganda, Nation Building, and Identity in Rwanda. *Journal of Political Economy*, 127, 1008-62.

Bonomi, G., Gennaioli, N. & Tabellini, G. (2021). Identity, Beliefs, and Political Conflict. *Quarterly Journal of Economics*, 136, 2371-2411.

Conley, T. (1999). GMM estimation with cross-sectional dependence. *Journal of Econometrics*, 91(1):1–45.

Dehdari, S. & Gehring, K. (2022). The Origins of Common Identity: Evidence from Alsace-Lorraine. *American Economic Journal: Applied*, 14, 261-92.

Dell, M. & Querubin, P. (2018). Nation Building through Foreign Intervention: Evidence from Discontinuities in Military Strategies. *Quarterly Journal of Economics*, 133, 701-64.

Dell, M. (2010). The Persistent Effects of Mining Mita. *Econometrica*, 78, 1863-1903.

Depetris-Chauvin, E., Durante, R., & Campante, F. (2020). Building Nations through Shared Experiences: Evidence from African Football. *American Economic Review*, 110, 1572-1602.

- Dincecco, M. & Onorato, M. (2018). *From Warfare to Wealth. The Military Origins of Urban Prosperity in Europe*. Boston: Cambridge University Press.
- Dincecco, M. & Prado, M. (2018). Warfare, fiscal capacity, and performance. *Journal of Economic Growth*, 17, 171-203.
- Dirección Nacional de Estadística (1933). *Extracto Estadístico y Censo Electoral de la República*. Lima: Dirección Nacional de Estadística.
- Dirección Nacional de Estadística y Censos (1965). *VI censo nacional de población*. Lima: Dirección Nacional de Estadística y Censos.
- Esposito, E., Rotesi, T., Saia, A. & Thoenig, M. (2021). *Reconciliation Narratives: The Birth of a Nation after the US Civil War*. CEPR Discussion Paper 15938.
- Feigenbaum, J., Lee, J., & Mezzanotti, F. (2022). Capital Destruction and Economic Growth: The Effects of Sherman's March, 1850-1920. *American Economic Journal: Applied*, 14, 301-42.
- Fontana, N., Nannicini, T. & Tabellini, G. (2017). Historical Roots of Political Extremism: The Effects of Nazi Occupation of Italy. *IZA Discussion Paper No. 10551*.
- Fouka, V. (2020). Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I. *Review of Economic Studies*, 87, 204-239.
- Gennaioli, N. & Voth, J. (2015). State Capacity and Military Conflict. *Review of Economic Studies*, 82, 1409-48.
- Gehring, K. (2022). Can external threats foster a European Union Identity? Evidence from Russia's Invasion of Ukraine. *The Economic Journal*, 132, 1489-1516.
- Giuliano, P. & Nunn, N. (2021). Understanding Cultural Persistence and Change. *Review of Economic Studies*, 88, 1541-1581.
- Grosjean, P. (2014). Conflict and Social and Political Preferences: Evidence from World War II and Civil Conflict in 35 European Countries. *Comparative Economic Studies*, 56, 424-451.

Guardado, J. (2018). Office-Selling, Corruption, and Long-Term Development in Peru. *American Political Science Review*, 112, 971-995.

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*, 20, 25–46.

Heldring, L., Robinson, J. & Whitfill, P. (2022). The Second World War, Inequality and The Social Contract in Britain. *NBER Working Paper 29677*.

Hobsbawm, E. J. (1990). *Nations and nationalism since 1780: Programme, myth, reality., The wiles lectures*. New York: Cambridge University Press

Huaroto, C. & Gallego, F. (2022). *The Legacy of the Spanish Conquista. Beliefs, Social Unrest, and Long-Term Economic Development*. Unpublished Manuscript.

Jurado Nacional de Elecciones (1982) Resultados de las Elecciones Políticas Generales de 1980. Lima: JNE. 2 v.20 cm.

Jha, S. (2013) Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia. *American Political Science Review*. 107, 806–832.

Kranton, R. (2016). Identity Economics 2016: Where do Social Distinctions and Norms com from? *American Economic Review: Papers & Proceedings*, 106, 405-409.

Lowes, S. & Montero, E. (2021a). Concessions, Violence, and Indirect Rule: Evidence from the Congo Free State. *Quarterly Journal of Economics*, 136 (4): 2047-91.

Lowes, S. & Montero, E. (2021b). The Legacy of Colonial Medicine in Central Africa. *American Economic Review*, 111: 1284-1314.

Lowes, S., Nunn, N., Robinson, J. & Weigel, J. (2017). The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom. *Econometrica*, 85, 1065-91.

Lupu, N. & Peisakhin, L. (2017). The Legacy of Political Violence across Generations. *American Journal of Political Science*, 61, 836-851.

Mallón, F. (1995). *Peasant and Nation. The Making of Postcolonial Mexico and Peru*. University of California Press.

Mangini, M. & Petroff, C. (2022). *Conflict Technology as a Catalyst of State Formation*. Unpublished manuscript.

Manrique, N. (1981). *Campesinado y Nación: Las Guerrillas Indígenas en la Guerra con Chile*. Lima: Centro de Investigación y Capacitación (CIC).

Marciante, G. (2022). *When Nation Building Goes South: Draft Evasion, Government Repression, and the Origins of the Sicilian Mafia*. Unpublished Manuscript.

Ministerio de Guerra (1981a). *La Guerra del Pacifico 1879-1883. La Resistencia de la Breña. Tomo I. De los reductos a Julcamarca*. Ministerio de Guerra: Lima.

Ministerio de Guerra (1981b). *La Guerra del Pacifico 1879-1883. La Resistencia de la Breña. Tomo II. La contraofensiva de 1882*. Ministerio de Guerra: Lima.

Ministerio de Guerra (1981c). *La Guerra del Pacifico 1879-1883. La Resistencia de la Breña. Tomo III. Huamachuco y el alma nacional (1882-1884)*. Ministerio de Guerra: Lima.

Ministerio de Hacienda y Comercio (1945). *Censo nacional de población de 1940*. Lima: Ministerio de Hacienda y Comercio, Dirección Nacional de Estadística y Censos.

Mitrunen, M. (2021). *Can You Make an American? Compulsory Patriotism and Assimilation of Immigrants*. Unpublished Manuscript.

Ministerio de Gobierno, Policía y Obras públicas (1878). *Censo general de la República del Perú, formado en 1876*. Lima: Imprenta del Teatro Nacional.

Ochsner, C. & Roesel, F. (2019). *Mobilizing History*. Unpublished Manuscript.

Oficina Nacional de Estadísticas y Censo (1974). *Censos nacionales VII de población, II de vivienda, 4 de junio de 1972: resultados definitivos, nivel nacional*. Lima: ONEC.

Oficina Nacional de Estadísticas y Censo (1968). *Censos nacionales de población, vivienda y agropecuario, 1961*. Lima: ONEC.

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics*, 37, 187-204.

- Ramankutty, N., Foley, J., Norman, J., & McSweeney, K. (2001). *Agricultural Suitability of Global Soils*. Dataset.
- Riaño, J. & Valencia, F. (2021). *Collateral Damage. The Legacy of the Secret War in Laos*. Unpublished manuscript.
- Rohner, D., Thoenig, M., & Zilibotti, F. (2013). Seeds of distrust: Conflict in Uganda. *Journal of Economic Growth*, 18, 217-252.
- Ronconi, J. & Ramos-Toro, D. (2022). *Nation Building Through Military Service*. Unpublished Manuscript.
- Sater, W. (2007). *Andean Tragedy: Fighting the War of the Pacific, 1879-1884*. Lincoln: Nebraska University Press.
- Shayo, M. (2009). A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution. *American Political Science Review*, 103: 147–74.
- Steiner, N., Berlinschi, R., Farvaque, E., Fidrmuc, J., Harms, P., Mihailov, A., Neugart, M., & Stanek, P. (2022). *Rallying around the EU Flag: Russia's Invasion of Ukraine and Attitudes toward European Integration*. CESIFO Working Papers 9883-2022.
- Teso, E. (2018). The Long-Term Effect of Demographic Shocks on the Evolution of Gender Roles: Evidence from the Transatlantic Slave Trade. *Journal of the European Economic Association*, 17, 497–534.
- Tilly, C. (1975). *The Formation of National States in Western Europe*. Princeton: Princeton University Press.
- Tur-Prats, A. & Valencia, F. (2020). *The Long Shadow of the Spanish Civil War*. CEPR Discussion Paper 15091.
- Voigtländer, N. & Voth, J. (2012) Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany. *Quarterly Journal of Economics*, 127, 1339-1392.

11. Figures and Tables

Figure 1

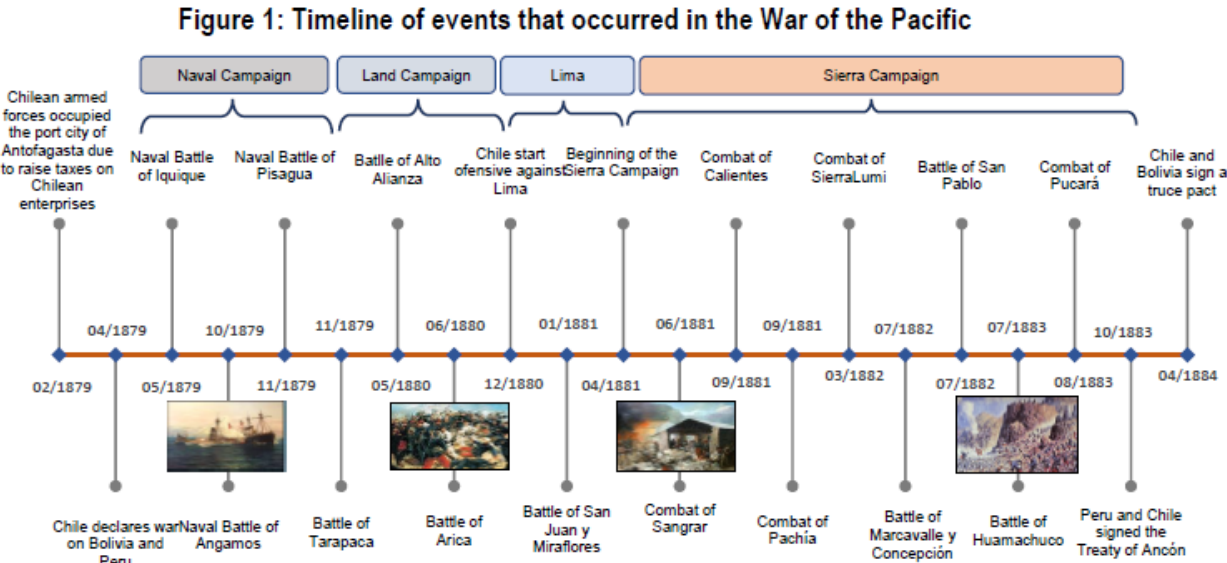


Figure 2

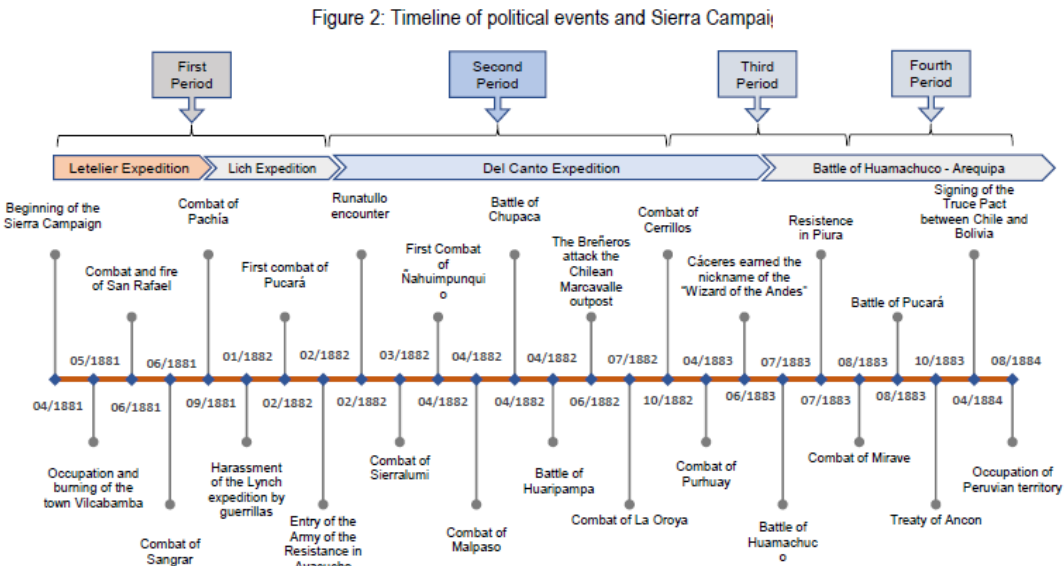
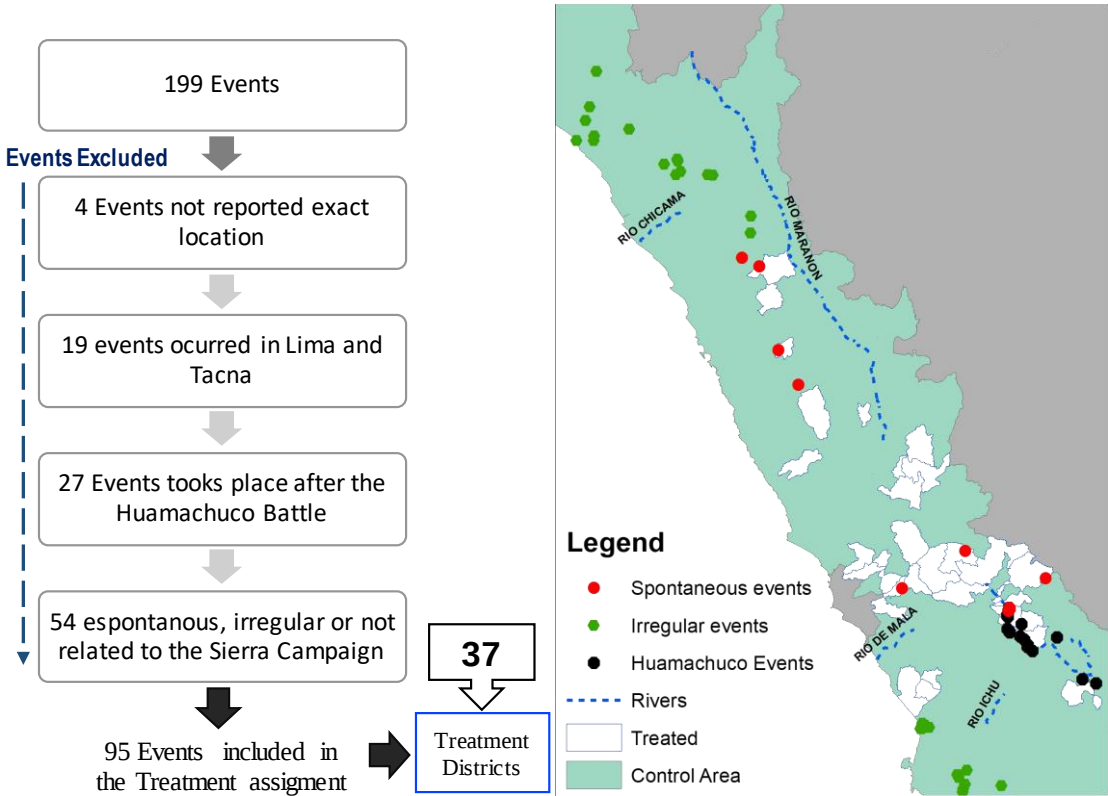


Figure 3: Definition of Treatment



1/ All events are defined as historical facts that occurred shortly before, during and after the Breña Campaign.

Figure 4. Location of the Sierra Campaign

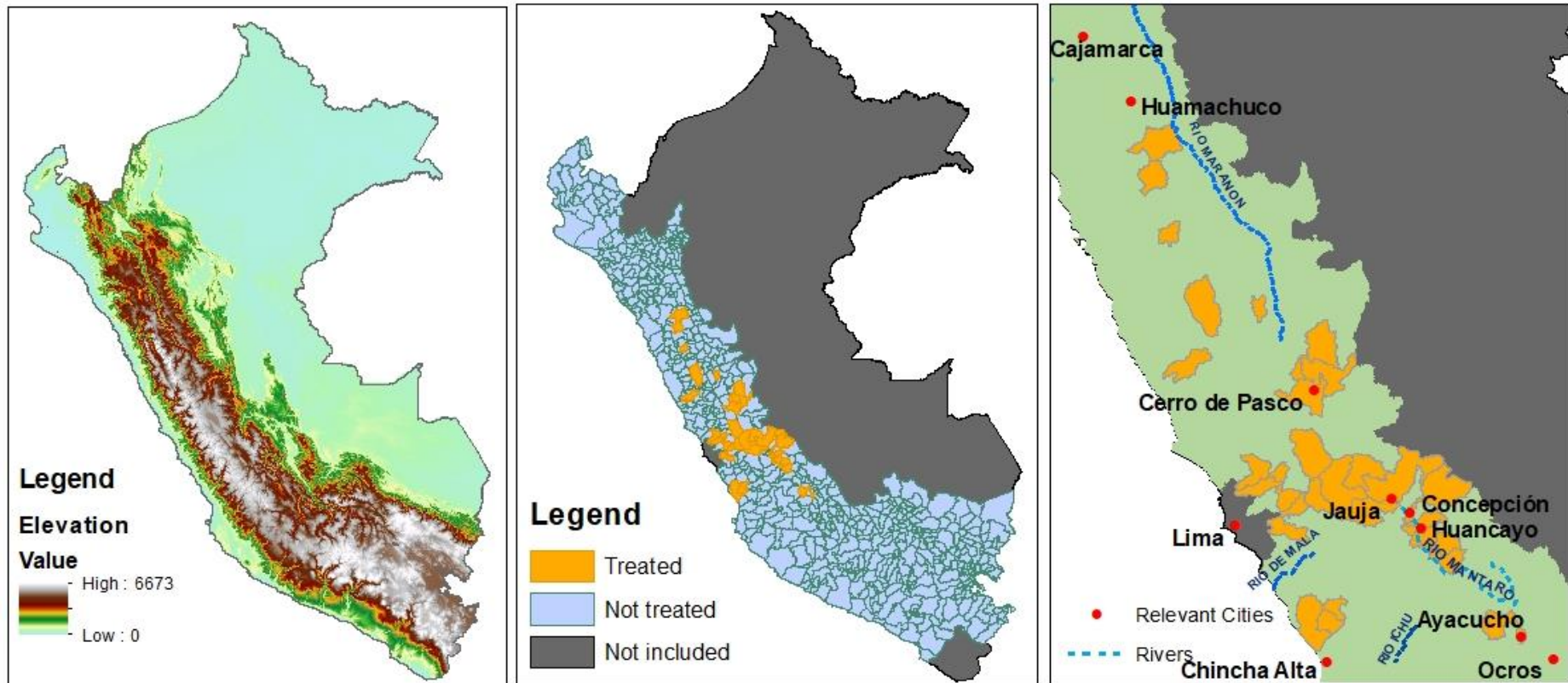


Figure 5: Heterogeneity of the effect on Identity by groups

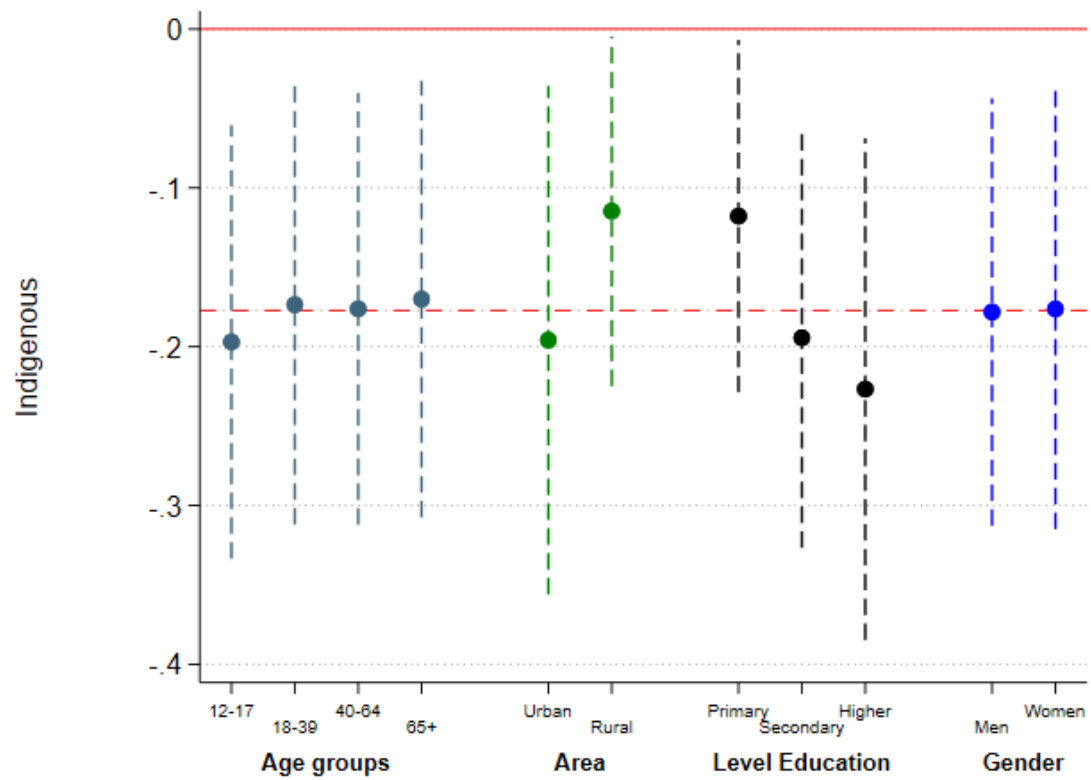


Table 1: Descriptive Statistics

Variable	N	Mean	S.D.	Min	Max	Source
<i>Treatment and geographic controls</i>						
Sierra Campaign (D = 1)	638	0.06	0.23	0.00	1.00	Own elaboration based on Ministerio de Guerra (1981a) (1981b) (1981c)
<i>Historical controls</i>						
Mita Districts (D = 1)	638	0.28	0.45	0.00	1.00	Own elaboration based on Dell (2010) and Huaroto and Gallego (2022)
Log of Pre-Colonial population in 1500	638	7.93	1.13	3.63	10.96	Own elaboration based on Golderwijk et al. (2010)
Share of Pre-Colonial cropland area in 1500	638	0.07	0.18	0.00	0.95	
<i>Geographical controls</i>						
Weighted altitude	638	2,988.4	1,329.1	0.0	4,772.1	Own elaboration based
Weighted slope	638	15.98	7.21	0.53	29.99	on SRTM data from NASA
Long-Term historical precipitation	638	5.64	3.92	0.05	41.02	Own elaboration based on Climate Research Unit data 1980-2017
Long-Term historical air temperature	638	10.62	6.03	-1.06	25.91	Own elaboration based on Agricultural Suitability of Global Soils data
Index of suitable land	638	0.29	0.37	0.00	0.99	
<i>1876 Controls</i>						
Log of Population Density	638	1.62	1.15	-1.52	6.92	Own elaboration based on MGPO (1878)
Percent of the rural population	638	0.50	0.31	0.00	0.99	
The ratio of men to women	638	1.04	0.37	0.00	4.48	
Log area	638	6.23	1.00	1.45	9.02	
Literacy rate	638	0.12	0.11	0.00	0.91	
Indigenous rate	638	0.01	0.00	0.00	0.01	
Percent of foreign population	638	0.02	0.07	0.00	0.63	
Percent of the catholic population	638	0.98	0.09	0.00	1.00	
Percent of people living on farms	638	0.15	0.19	0.00	0.94	
<i>Table 2, 3, 4 & 5: Effect on Identity Beliefs</i>						
Percent of people identify as indigenous	12,175,049	0.32	0.47	0.00	1.00	Microdata from the Nacional Census 2017 (INEI)
Percent of people identify as white	12,175,049	0.06	0.23	0.00	1.00	
Percent of people identify as mestizo	12,175,049	0.54	0.50	0.00	1.00	
Percent of people identify as others	12,175,049	0.08	0.27	0.00	1.00	
Percent of people identify as indigenous	53,827	0.34	0.47	0.00	1.00	Microdata from the National Household Survey 2017 (INEI)
Percent of people identify as white	53,827	0.05	0.21	0.00	1.00	

Table 1: Descriptive Statistics

Variable	N	Mean	S.D.	Min	Max	Source
Percent of people identify as mestizo	53,827	0.45	0.50	0.00	1.00	
Percent of people identify as others	53,827	0.16	0.37	0.00	1.00	

Table 6: Effect on First Learned Language or Mother Tongue (MT)

Share of people with <i>Spanish</i> as MT in 1961	638	0.45	0.43	0.00	1.00	<i>Own elaboration based on DNEC (1965)</i>
Share of people with <i>Native</i> as MT in 1961	638	0.54	0.43	0.00	0.99	
Share of people with <i>Others</i> MT in 1961	638	0.01	0.01	0.00	0.05	
Share of people with <i>Spanish</i> as MT in 1981	608	0.55	0.41	0.00	1.00	<i>Microdata from National Census 1981 (INEI)</i>
Share of people with <i>Native</i> as MT in 1981	608	0.45	0.41	0.00	1.00	
Share of people with <i>Others</i> MT in 1981	608	0.00	0.01	0.00	0.09	
Share of people with <i>Spanish</i> as MT in 2017	638	0.61	0.36	0.01	1.00	<i>Microdata from National Census 2017 (INEI)</i>
Share of people with <i>Native</i> as MT in 2017	638	0.38	0.36	0.00	0.98	
Share of people with <i>Others</i> MT in 2017	638	0.01	0.02	0.00	0.14	

Table 7: Effect on Indigenous Rate

The Indigenous rate at 1876	638	0.61	0.27	0.00	1.00	<i>Own elaboration based on MGPO (1878)</i>
The Indigenous rate at 1940	638	0.50	0.33	0.00	0.99	<i>Own elaboration based on MHC (1945)</i>

Table 8: Effect on Presidential Elections

Log of registered pop in 1933	634	5.48	1.13	0.00	8.94	<i>Own elaboration based on DNE (1933)</i>
Log of voters in the election at 1933	634	6.37	1.05	3.37	9.93	
Log of registered pop in 1980	634	7.21	1.20	4.16	11.63	<i>Own elaboration based on JNE (1980)</i>
Log of voters in the election at 1980	634	8.28	1.10	4.80	12.23	
Log of registered pop in 2001	634	7.92	1.37	4.85	12.68	<i>Microdata from ONPE</i>
Log of voters in the election at 2001	634	8.45	1.28	5.50	12.95	
Log of registered pop in 2021	634	8.19	1.37	5.16	12.96	
Log of voters in the election at 2021	634	8.89	1.30	5.89	13.46	

Table 9: Effect on Civic Education

Z-score achieved in civic education	5,200	0.03	1.00	-3.58	3.97	<i>Microdata from the National Student Assessment of the</i>
Z-score achieved in math	5,433	0.02	1.01	-4.88	4.84	
Z-score achieved in reading	5,434	0.01	1.00	-5.26	4.27	

Table 1: Descriptive Statistics

Variable	N	Mean	S.D.	Min	Max	Source
						Ministry of Education (2018)
Table 10: Effect on Democracy and Self-Identification						
Knows what democracy is	12,844	0.40	0.49	0.00	1.00	Microdata from the National Household Survey 2017 (INEI)
Democracy is important for: Choosing authorities	5,200	0.38	0.48	0.00	1.00	
Democracy is important for: Being represented	5,200	0.22	0.42	0.00	1.00	
Do you identify more with Country, Region, or Province	12,844	0.57	0.50	0.00	1.00	
Do you identify more with Peasant Community	12,844	0.20	0.40	0.00	1.00	
Table 11: Effect on Literacy Rate						
Percent of Literacy rate in 1940	638	0.27	0.18	0.03	0.72	Own elaboration base don MHC (1945) and DNEC (1965)
Percent of Literacy rate in 1961	638	0.48	0.22	0.00	0.94	
Percent of Literacy rate in 1981	638	0.57	0.13	0.22	0.86	Microdata from National Census (INEI)
Percent of Literacy rate in 1993	638	0.65	0.11	0.32	0.89	
Table 12: Effect on House Quality & Welfare						
Share of houses with good wall quality in 1961	638	0.72	0.24	0.01	1.00	Own elaboration based on DNEC (1965)
Share of houses with good roof quality in 1961	638	0.37	0.28	0.00	0.99	
Share of houses with good floor quality in 1961	638	0.11	0.16	0.00	0.88	
Houses with dirt-floor	4,192,391	0.44	0.50	0.00	1.00	Microdata from National Census (INEI)
Log of per capita income 2017	21,907	8.72	0.83	4.89	12.11	Microdata from National Household Survey (INEI)

Table 2. Long-Term Effect of the Sierra Campaign on Ethnic Identity

	Q: To which of the following groups do you identify the most?							
	Population Census 2017				ENAH0 2021 Survey			
	Indigenous	White	Mestizo	Other ^{1/}	Indigenous	White	Mestizo	Other ^{1/}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sierra Campaign	-0.17** (0.07)	0.01 (0.01)	0.17*** (0.06)	-0.003 (0.01)	-0.14** (0.07)	-0.001 (0.01)	0.16*** (0.05)	-0.01 (0.03)
R-squared	0.37	0.02	0.21	0.02	0.43	0.03	0.20	0.12
N	12,159,826	12,159,826	12,159,826	12,159,826	53,823	53,823	53,823	53,823
Mean Control	0.31	0.06	0.55	0.08	0.33	0.05	0.44	0.18
1876 Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Historical Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic and Climatic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

The Table reports OLS estimates. Data is at the individual level. The dependent variables for identity beliefs were calculated from the respondent's answer to the question: To which groups do you identify the most? ^{1/}Others include Afrodescendants and other identity responses. The complete list of included control variables is presented in Table 1. Estimates include control of individuals such as gender, age, area, and education level. Standard errors are clustered at the district level. Coefficients significantly different from zero are denoted by: *** p< 0.01, ** p<0.05, *p<0.1.

Table 3. Long-Term Effect of the Sierra Campaign on Ethnic Identity with Bad Controls

Q: To which of the following groups do you identify the most?								
	Population Census 2017				ENAH0 2017 Survey			
	Bad control: Dirt-floor House				Bad control: Per cápita Household income			
	Indigenous (1)	White (2)	Mestizo (3)	Other ^{1/} (4)	Indigenous (5)	White (6)	Mestizo (7)	Other ^{1/} (8)
Sierra Campaign	-0.18** (0.07)	0.01 (0.01)	0.17*** (0.06)	-0.002 (0.01)	-0.14** (0.07)	-0.001 (0.01)	0.16*** (0.05)	-0.01 (0.03)
Bad Controls	0.04*** (0.01)	-0.002** (0.001)	-0.04*** (0.01)	0.01** (0.003)	-0.01** (0.01)	-0.002 (0.003)	0.03*** (0.01)	-0.02*** (0.01)
R-squared	0.39	0.02	0.21	0.03	0.43	0.03	0.20	0.13
N	11,707,824	11,707,824	11,707,824	11,707,824	53,823	53,823	53,823	53,823
Mean Control	0.31	0.06	0.55	0.08	0.33	0.05	0.44	0.18
1876 Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Historical Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic and Climatic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

The Table reports OLS estimates. Data is at the individual level. The dependent variables for identity beliefs were calculated from the respondent's answer to the question: To which groups do you identify the most? ^{1/}Others include Afrodescendants and other identity responses. The complete list of included control variables is presented in Table 1. Estimates include control of individuals such as gender, age, area, and education level. Standard errors are clustered at the district level. Coefficients significantly different from zero are denoted by: *** p< 0.01, ** p<0.05, *p<0.1.

Table 4. Long-Term Effect of the Sierra Campaign on Ethnic Identity by type of resident (Native or Migrant)

Q: To which of the following groups do you identify the most?								
	Population Census 2017				ENAH0 2017 Survey			
	Indigenous	White	Mestizo	Other	Indigenous	White	Mestizo	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Native	-0.2*** (0.06)	0.01 (0.01)	0.18*** (0.05)	0.004 (0.01)	-0.17*** (0.06)	0.001 (0.01)	0.18*** (0.06)	-0.004 (0.03)
Migrant	-0.04 (0.07)	0 (0.01)	0.06 (0.06)	-0.03*** (0.01)	0.07 (0.06)	-0.01 (0.01)	-0.03 (0.05)	-0.03 (0.03)
R-squared	0.37	0.02	0.21	0.02	0.43	0.03	0.20	0.12
N	12,160,806	12,160,806	12,160,806	12,160,806	53,823	53,823	53,823	53,823
Mean Control (MC)	0.31	0.06	0.55	0.08	0.33	0.05	0.45	0.18
1876 Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Historical Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic and Climatic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

The Table reports OLS estimates. Data is at the individual level. The dependent variables for identity beliefs were calculated from the respondent's answer to the question: To which groups do you identify the most? ^{1/}Others include Afrodescendants and other identity responses. The complete list of included control variables is presented in Table 1. Estimates include control of individuals such as gender, age, area, and education level. Standard errors are clustered at the district level. Coefficients significantly different from zero are denoted by: *** p< 0.01, ** p<0.05, *p<0.1.

Table 5. Long-Term Effect of the Sierra Campaign on Identity beliefs of Migrants living in Lima city

Q: To which of the following groups do you identify the most?								
	Population Census 2017				ENAH0 2017 Survey			
	Indigenous	White	Mestizo	Other ^{1/}	Indigenous	White	Mestizo	Other ^{1/}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sierra Campaign	-0.1*** (0.03)	-0.002 (0.004)	0.09*** (0.03)	0.004 (0.003)	-0.13*** (0.03)	0 (0.01)	0.15*** (0.03)	-0.02 (0.02)
R-squared	0.18	0.01	0.11	0.01	0.26	0.04	0.17	0.09
N	2,273,367	2,273,367	2,273,367	2,273,367	2,522	2,522	2,522	2,522
Mean Control	0.27	0.06	0.62	0.06	0.34	0.05	0.45	0.17
1876 Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Historical Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic and Climatic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table reports OLS estimates. Data is at the individual level. The dependent variables for identity beliefs were calculated from the respondent's answer to the question: To which groups do you identify the most? ^{1/}Others include Afrodescendants and other identity responses. The complete list of included control variables is presented in Table 1. Estimates include control of individuals such as gender, age, area, and education level. Standard errors are clustered at the district level. Coefficients significantly different from zero are denoted by: *** p< 0.01, ** p<0.05, *p<0.1.

Table 6. Long-Term Effect of the Sierra Campaign on First Learned Language or Mother Tongue

Which language did you learn in your childhood?									
	1961			1981			2017		
	Spanish	Indigenous	Other ^{1/}	Spanish	Indigenous	Other ^{1/}	Spanish	Indigenous	Other ^{1/}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sierra Campaign	9.71** (4.42)	-9.96** (4.46)	0.25*** (0.09)	23.69*** (6.44)	-23.53*** (6.55)	-0.15 (0.25)	22.73*** (6.94)	-23.04*** (6.82)	0.31 (0.33)
R-squared	0.72	0.72	0.19	0.68	0.69	0.45	0.64	0.64	0.25
N	638	638	638	608	608	608	638	638	638
N - expand	5,794,513	5,794,513	5,794,513	9,755,880	9,755,880	9,755,880	15,551,085	15,551,085	15,551,085
Mean Control	45.35	53.90	0.75	53.63	45.99	0.38	59.54	39.17	1.29
1876 Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Historical Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic and Climatic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table reports OLS estimates. Data is at the district level using the population in the district as an expansion factor. The dependent variables for identity beliefs were calculated from the respondent's answer to the question: To which groups do you identify the most? ^{1/}Others include Afrodescendants and other identity responses. The complete list of included control variables is presented in Table 1. Robust standard errors are reported. Coefficients significantly different from zero are denoted by: *** p< 0.01, ** p<0.05, *p<0.1.

Table 7. Effect of the Sierra Campaign on Race in 1876 and 1940

	Percentage of the population that is indigenous			
	1876		1940	
	(1)	(2)	(3)	(4)
Sierra Campaign	1.82 (3.55)	0.92 (3.72)	-1.72 (2.7)	-2.99 (2.72)
R-squared	0.36	0.42	0.67	0.70
N	638	638	638	638
N - expand	2,288,491	2,288,491	5,009,001	5,009,001
Mean Control	62.70	62.70	51.23	51.23
1876 Controls	Yes	Yes	Yes	Yes
Historical control	Yes	Yes	Yes	Yes
Geographic and Climatic Controls	No	Yes	No	Yes

Notes: Table reports OLS estimates. Data is at the district level using the population in the district as an expansion factor. The dependent variables for identity beliefs were calculated from the respondent's answer to the question: To which groups do you identify the most? ¹Others include Afrodescendants and other identity responses. The complete list of included control variables is presented in Table 1. Robust standard errors are reported. Coefficients significantly different from zero are denoted by: *** p< 0.01, ** p<0.05, *p<0.1.

Table 8. Effect of the Sierra Campaign on Electoral Participation in 1933, 1981, 2001 and 2021

	Restricted elections		Mandatory and elections					
	Presidential 1933		Presidential 1980		Presidential 2001		Presidential 2021	
	Log of registered pop		Log of registered pop		Log of registered pop		Log of registered pop	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sierra Campaign (D=1)	0.19*** (0.06)	0.19*** (0.06)	0.19*** (0.06)	0.06 (0.06)	0.11*** (0.02)	0.1*** (0.02)	0.07*** (0.02)	0.06*** (0.02)
Voters in elections (logs)	0.87*** (0.03)	0.28** (0.12)	0.94*** (0.03)	0.03 (0.17)	1.05*** (0.01)	0.95*** (0.03)	1.06*** (0.01)	0.96*** (0.04)
Rural population (logs)	-0.05* (0.02)	-0.04* (0.02)	-0.04* (0.02)	-0.03* (0.02)	-0.01 (0)	-0.01 (0.01)	-0.02*** (0.01)	-0.02*** (0.01)
Literacy population (logs)		0.58*** (0.11)		0.87*** (0.16)		0.11*** (0.03)		0.1*** (0.04)
R^2	0.84	0.84	0.90	0.92	0.99	0.99	0.99	0.99
N	634	634	634	634	634	634	634	634
Mean Control	5.4	5.4	7.2	7.2	7.9	7.9	8.2	8.2
1876 Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Historical Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic and Climatic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table reports OLS estimates. Data is at the district level. The dependent variables for identity beliefs were calculated from the respondent's answer to the question: To which groups do you identify the most? ¹Others include Afrodescendants and other identity responses. The complete list of included control variables is presented in Table 1. Robust standard errors are reported. Coefficients significantly different from zero are denoted by: *** p< 0.01, ** p<0.05, *p<0.1.

Table 9. Long-Term Effect of the Sierra Campaign on Civic Education

Dependent: Z-score achieved			
	Civic	Maths	Reading
	(1)	(2)	(3)
Sierra Campaign	0.26** (0.12)	0.14 (0.16)	0.14 (0.13)
R2	0.17	0.08	0.17
N	5,200	5,433	5,434
1876 Controls	Yes	Yes	Yes
Historical Control	Yes	Yes	Yes
Geographic and Climatic Controls	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
School-level Controls	Yes	Yes	Yes

Notes: Table reports OLS estimates. Data is at the student level. The control variables are described in Table 1. Individual controls include a control for the student's gender and age. Standard errors are clustered at the district level. School controls include the gender of students and dummies by type, area, and management. The dependent variable is the Z-score achieved in the civic education subject of 6th grade and math and reading of 2nd-grade elementary school students. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 10. Long-Term Effect of the Sierra Campaign on Beliefs around Democracy and Self-Identification with the country

	Belief around Democracy	Democracy is important for:		Do you identify more with your ...?	
	Knows what democracy is	Choose authorities	To be represented	Country, Region, or Province	Peasant Community
	(1)	(2)	(3)	(4)	(5)
Urban	0.02 (0.03)	0.18* (0.1)	0.206 (0.13)	0.055 (0.04)	-0.03 (0.03)
Rural	0.11*** (0.04)	0.25*** (0.09)	0.22** (0.1)	0.151*** (0.04)	-0.11*** (0.04)
R ²	0.12	0.04	0.06	0.05	0.17
N	12,844	5,200	5,200	12,844	12,844
Control Mean	0.40	0.35	0.19	0.56	0.20
1876 Controls	Yes	Yes	Yes	Yes	Yes
Historical Control	Yes	Yes	Yes	Yes	Yes
Geographic and Climatic Controls	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes

Notes: Table reports OLS estimates. Data is at the individual level. The dependent variables come from three questions. The first is "Do you know what democracy is?" for which I use a dummy if there is a positive answer. The second is asked only to the ones that previous: "What is democracy for?", for which they have to select all alternatives that they consider valid from a list of five: i) Elect authorities, ii) Be represented, iii) Achieve family wellbeing, iv) Make people's right respected, and v) For nothing. The third was "to which group do you identify the most?" for which they had to select an alternative from i) Region, province, district, or city, ii) Ethnic group or race, iii) Indigenous peasant community, iv) Religious group or v) Others. The complete list of included control variables is presented in Table 1. Estimates include control of individuals such as gender, age, area, and education level. Standard errors are clustered at the district level. Coefficients significantly different from zero are denoted by: *** p< 0.01, ** p<0.05, *p<0.1.

Table 11. Diff and Diff Estimates of the Effect of the *Sierra Campaign* on Literacy

	1940	1961	1981	1993
	(1)	(2)	(3)	(4)
Panel A: Benchmark				
Sierra Campaign*Post	7.91*** (2.16)	12.2*** (2.86)	8.86*** (1.71)	7.08*** (1.56)
Rurality rate 1876*SC*Post	17.45** (8.89)	18.33 (11.2)	10.64 (7.6)	14.61** (7.23)
R ²	0.93	0.95	0.97	0.98
Panel B: Including all sets of controls				
Sierra Campaign*Post	7.97*** (2.07)	12.81*** (2.62)	8.33*** (1.61)	6.99*** (1.45)
Rurality rate 1876*SC*Post	19.92** (8.64)	22.66** (10.44)	12.82* (7.25)	15.19** (6.78)
R ²	0.96	0.97	0.98	0.99
N	1,276	1,276	1,276	1,276
N - expand	7,300,879	9,282,579	12,371,104	14,840,545
Mean Control in 1876	12.21	12.21	12.21	12.21
Mean Control in each year	26.31	47.22	56.03	64.49
District Fixed Effects	Yes	Yes	Yes	Yes
1876 Controls*Post	Yes	Yes	Yes	Yes
Historical Controls*Post	Yes	Yes	Yes	Yes
Geographic and Climatic Controls*Post	Yes	Yes	Yes	Yes

Notes: Tables report DID estimates. Data is at the district-year level and uses population as an expansion factor. Post is a dummy variable that takes the value of one if it's after the SC. The dependent variable is the literacy rate, and the interaction term is the percentage of the population in 1876 that lived in rural areas. The complete list of included control variables is presented in Table 1. Robust standard errors are shown. Coefficients significantly different from zero are denoted by: *** p< 0.01, ** p<0.05, *p<0.1.

Table 12. OLS Estimates of the Effect of the Sierra Campaign on Economic Development proxies

	House quality 1961			Welfare 2017	
	Good wall quality ^{1/}	Good roof quality ^{2/}	Good floor quality ^{3/}	Dirt- Floor	Log. of the HH Income per capita
	(1)	(2)	(3)	(4)	(5)
Sierra Campaign	0.02 (0.02)	0.21*** (0.04)	0.09*** (0.03)	-0.19*** (0.06)	0.17* (0.1)
R-squared	0.46	0.40	0.50	0.22	0.23
N	638	638	638	4,192,391	21,907
Mean Control	0.71	0.35	0.10	0.45	8.71
1876 Controls	Yes	Yes	Yes	Yes	Yes
Historical Control	Yes	Yes	Yes	Yes	Yes
Geographic and Climatic Controls	Yes	Yes	Yes	Yes	Yes

Notes: Table reports OLS estimates. Data for columns 1 to 3 is at the district level, and data for columns 4 and 5 are at the individual level. The first three dependent variables are the share of houses in each district with good-quality walls, roofs, and floor materials. The complete list of included control variables is presented in Table 1. Standard errors are clustered at the district level. ^{1/} Bricks, concrete, blocks, stone, or cooked mud (adobe). ^{2/} Cement, clay tiles, plates, calamine, wood, or similar. ^{3/} Wood, concrete, and any other material different from dirt. significantly different from zero are denoted by: *** p< 0.01, ** p<0.05, *p<0.1.

12. Appendix Figures and Tables

Figure A1: Controls Regression

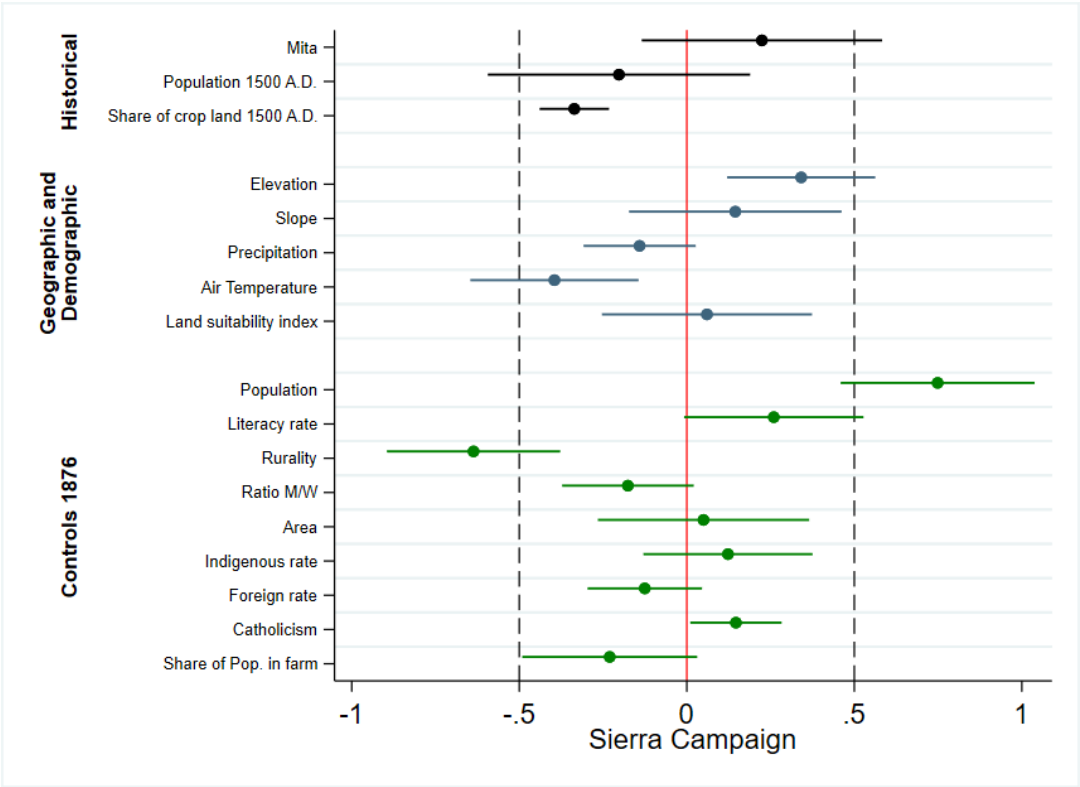


Figure A2: Location of Control group by different matching algorithms

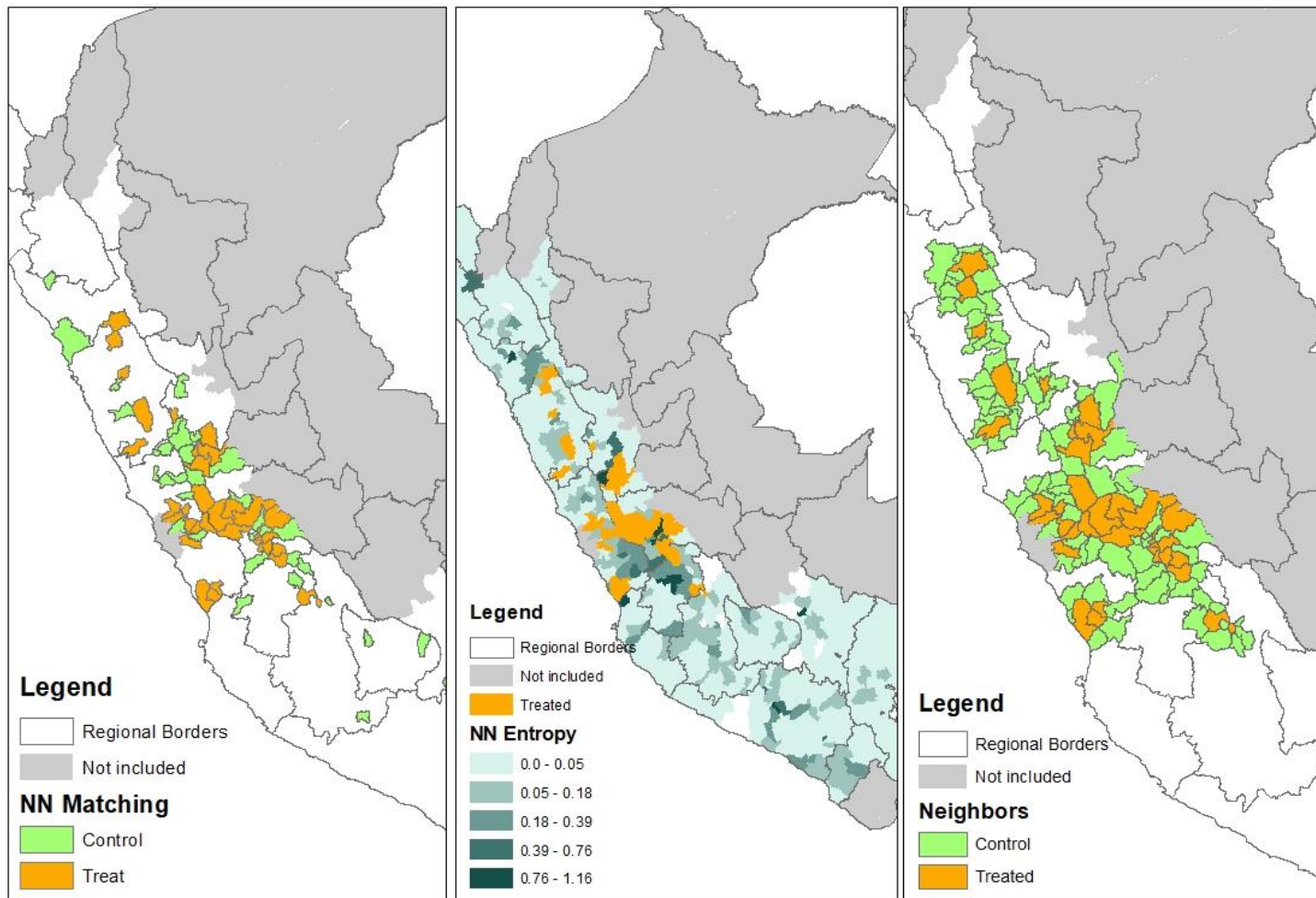


Figure A3: Randomization Inference Robustness Check of results from Table 2

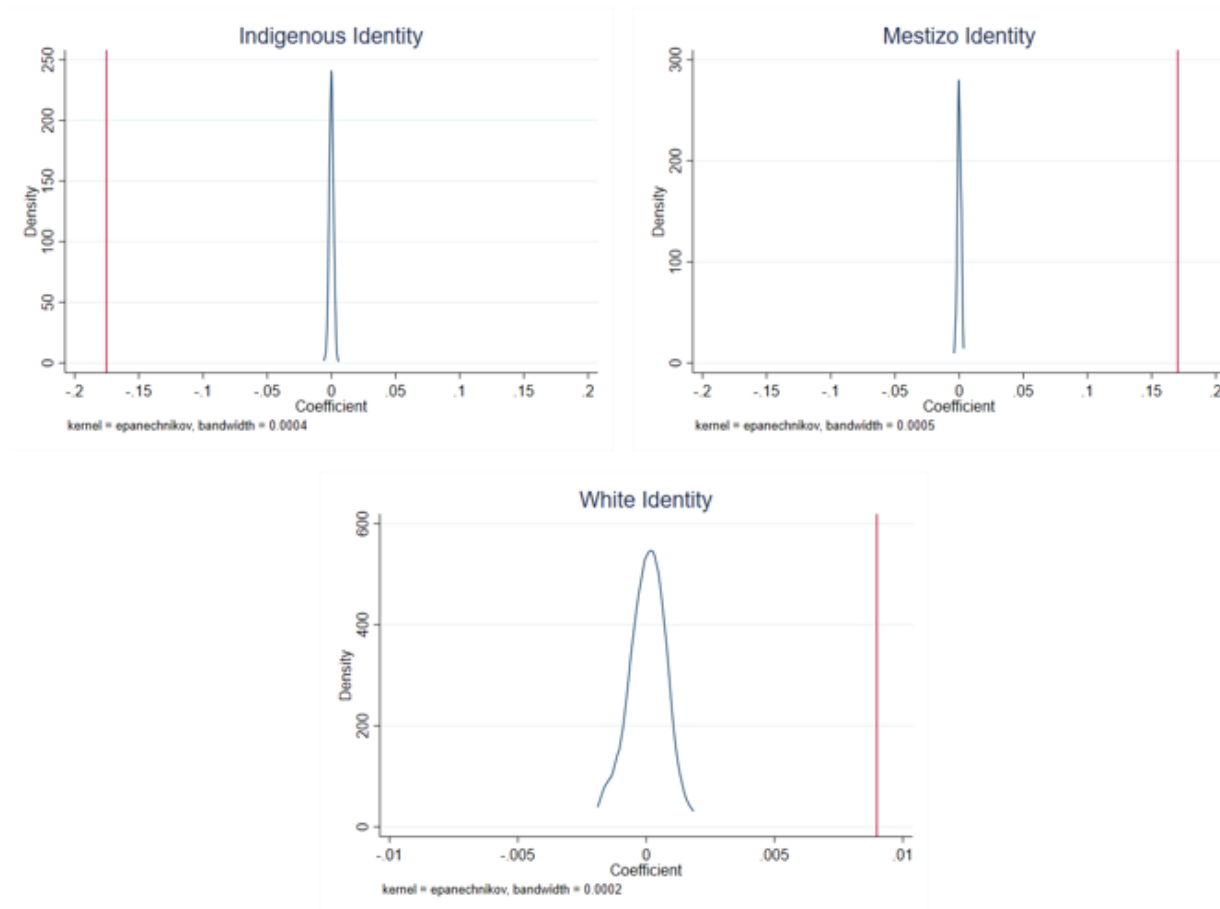


Table A.1. Long-Term Effect of the Sierra Campaign on Turnout and Valid Vote rate

	Share of people that...in the elections			
	Presidential Elections in 2021		Municipal and Regional Elections in 2018	
	Turn Out	% Valid Vote	Turn Out	% Valid Vote
	(1)	(2)	(3)	(4)
Sierra Campaign	2.19* (1.22)	3.05*** (0.69)	8.9*** (1.78)	8.26*** (2.76)
R-squared	0.44	0.41	0.09	0.09
N	638	638	638	638
Mean Control	67.22	75.32	75.58	81.25
1876 Controls	Yes	Yes	Yes	Yes
Historical Control	Yes	Yes	Yes	Yes
Geographic and Climatic Controls	Yes	Yes	Yes	Yes

Notes: Table reports OLS estimates. Data is at the district level. The dependent variables for identity beliefs were calculated from the respondent's answer to the question: To which groups do you identify the most? 1/Others include Afrodescendants and other identity responses. The complete list of included control variables is presented in Table 1. Robust standard errors are reported. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.2. Identity Effect of the Sierra Campaign: Matching in Observables

	Benchmark	NN Matching	Entropy Balance	Neighbors
	(1)	(2)	(4)	(3)
Panel A: Indigenous (Quechua, Aymara, ect)				
Sierra Campaign	-0.17** (0.07)	-0.1* (0.06)	-0.11*** (0.04)	-0.06* (0.03)
R^2	0.37	0.21	0.19	0.21
N	12,159,826	1,490,419	12,159,826	2,352,657
Panel B: Mestizo (Cholo, trigueño, etc)				
Sierra Campaign	0.17*** (0.06)	0.1* (0.06)	0.1*** (0.03)	0.05* (0.03)
R^2	0.21	0.15	0.13	0.15
N	12,159,826	1,490,419	12,159,826	2,352,657
Panel C: White				
Sierra Campaign	0.01 (0.01)	0.01*** (0)	0.01** (0)	0.01*** (0)
R^2	0.02	0.01	0.01	0.01
N	12,159,826	1,490,419	12,159,826	2,352,657
1876 Controls	Yes	Yes	Yes	Yes
Historical Control	Yes	Yes	Yes	Yes
Geographic and Climatic Controls	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes

Notes: Table reports OLS estimates after matching. Data is at the individual level. The dependent variables for identity beliefs were calculated from the respondent's answer to the question: To which groups do you identify the most? Estimates include control of individuals such as gender, age, area, and education level. The complete list of included control variables is presented in Table 1. Errors are clustered at the district level. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.3: Robustness for Unobservable Selection

To which of the following groups
do you identify the most?

	Indigenous		White		Mestizo		Others	
	no controls	with controls	no controls	with controls	no controls	with controls	no controls	with controls
	(1)	(2)	(3)	(4)	(5)	(6)	(5)	(6)
Sierra Campaign	0.12* (0.07)	-0.18** (0.07)	-0.02*** (0.01)	0.01 (0.01)	-0.07 (0.06)	0.17*** (0.06)	-0.04*** (0.01)	-0.002 (0.01)
N	12,175,049	12,175,049	12,175,049	12,175,049	12,175,049	12,175,049	12,175,049	12,175,049
R2	0.01	0.37	0.00	0.01	0.00	0.20	0.00	0.02
δ		-1.527		-0.766		-1.854		0.147
Identified Set		[-0.3, -0.18]		[0.01, 0.02]		[0.17, 0.27]		[-0.002, 0.01]

Notes: Table reports OLS estimates. Columns 1, 3, 5, and 7 report estimates with no control variables, only the treatment variable dummy. Columns 2,4,6, and 8 with all control variables used in Table 2. The δ row shows the amount of selection on unobservables necessary, relative to the degree of selection on observable controls, to explain away the coefficient in the respective column. For the calculation of this δ , I use the Stata command -psacalc-, setting Rmax to 1.3 times the R2 in the respective column; for details, see text and Oster (2017). The identified set in the final row is bound below β' and above β^* , calculated based on Rmax and $\delta=1$. Coefficients significantly different from zero are denoted by: y: *** p< 0.01, ** p<0.05, *. Robust standard errors.

Table A.4. The Effect of the Sierra Campaign on Ethnic Identity using Conley Standard errors

	To which of the following groups do you identify the most?			
	Indigenous	White	Mestizo	Other ^{1/}
	(1)	(2)	(3)	(4)
Sierra Campaign	-0.17** (0.07)	0.01 (0.01)	0.17*** (0.06)	0 (0.01)
R-squared	0.37	0.02	0.21	0.02
N	12,159,826	12,159,826	12,159,826	12,159,826
Conley standard errors in brackets				
(<50 km)	[0.05]***	[0.01]	[0.05]***	[0.01]
(<100 km)	[0.06]***	[0.01]	[0.05]***	<0.01>
(<150 km)	[0.05]***	[0.01]	[0.04]***	[0]*
Mean Control	0.31	0.06	0.55	0.08
1876 Controls	Yes	Yes	Yes	Yes
Historical Controls	Yes	Yes	Yes	Yes
Geographic and Climatic Controls	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes

Notes: Table reports OLS estimates. Data is at the individual level. The dependent variables for identity beliefs were calculated from the respondent's answer to the question: "To which groups do you identify the most?" ^{1/} Others include Afrodescendants and other identity responses. The complete list of included control variables is presented in Table 1. Estimates include control of individuals such as gender, age, area, and education level. Standard errors are clustered at the district level. Coefficients significantly different from zero are denoted by: *** p< 0.01, ** p<0.05, *p<0.1.